

Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach



Ravinesh C. Deo^{a,*}, Mehmet Şahin^b, Jan F. Adamowski^c, Jianchun Mi^{d,*}

^a School of Agricultural Computational and Environmental Sciences, Centre for Sustainable Agricultural Systems & Centre for Applied Climate Sciences, Institute of Life Sciences and the Environment, University of Southern Queensland, QLD 4300, Australia

^b Department of Electrical and Electronics Engineering, Siirt University, 56100 Siirt, Turkey

^c Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, Canada

^d Department of Energy & Resources, College of Engineering, Peking University, Beijing, China

ARTICLE INFO

Keywords:

Satellite solar model
Remote sensing
Extreme learning machine
Spatial forecasting

ABSTRACT

Global advocacy to mitigate climate change impacts on pristine environments, wildlife, ecology, and health has led scientists to design technologies that harness solar energy with remotely sensed, freely available data. This paper presents a study that designed a regionally adaptable and predictively efficient extreme learning machine (ELM) model to forecast long-term incident solar radiation (*ISR*) over Australia. The relevant satellite-based input data extracted from the Moderate Resolution Imaging Spectroradiometer (*i.e.*, normalized vegetation index, land-surface temperature, cloud top pressure, cloud top temperature, cloud effective emissivity, cloud height, ozone and near infrared-clear water vapour), enriched by geo-temporal input variables (*i.e.*, periodicity, latitude, longitude and elevation) are applied for a total of 41 study sites distributed approximately uniformly and paired with ground-based *ISR* (target). Of the 41 sites, 26 are incorporated in an ELM algorithm for the design of a universal model, and the remainder are used for model cross-validation. A universally-trained ELM (with training data as a global input matrix) is constructed, and the spatially-deployable model is applied at 15 test sites. The optimal ELM model is attained by trial and error to optimize the hidden layer activation functions for feature extraction and is benchmarked with competitive artificial intelligence algorithms: random forest (RF), M5 Tree, and multivariate adaptive regression spline (MARS). Statistical metrics show that the universally-trained ELM model has very good accuracy and outperforms RF, M5 Tree, and MARS models. With a distinct geographic signature, the ELM model registers a Legates & McCabe's Index of 0.555–0.896 vs. 0.411–0.858 (RF), 0.434–0.811 (M5 Tree), and 0.113–0.868 (MARS). The relative root-mean-square (RMS) error of ELM is low, ranging from approximately 3.715–7.191% vs. 4.907–10.784% (RF), 7.111–11.169% (M5 Tree) and 4.591–18.344% (MARS). Taylor diagrams that illustrate model precision in terms of RMS centred difference, error analysis, and boxplots of forecasted vs. observed *ISR* also confirmed the versatility of the ELM in generating forecasts over heterogeneous, remote spatial sites. This study ascertains that the proposed methodology has practical implications for regional energy modelling, particularly at national scales by utilizing remotely-sensed satellite data, and thus, may be useful for energy feasibility studies at future solar-powered sites. The approach is

List of Acronyms: ELM, Extreme Learning Machine; MARS, Multivariate Adaptive regressions Splines; RF, Random Forest; ISR, Incident Solar Radiation; NEM, National Electricity Market; AERA, Australian Renewable Energy Assessment; ANN, Artificial Neural Network; NOAA, National Oceanic and Atmospheric Administration; AR, autoregressive; AVHRR, Advanced Very High Resolution Radiometer; MLR, Multiple Linear Regressions; MERRA, Modern Era Retrospective-Analysis for Research & Applications; SVM, Support Vector Machine; CARDS, Coupled Auto-Regressive and Dynamical System; *N*, Length of dataset; MODIS, Moderate Resolution Imaging Spectroradiometer; (*X*, *Y*), Matrix of Input & Target; ARIMA, Autoregressive Integrated Moving Average; *p*, Reflectance ratio; W-SVM, Support Vector Machine-Wavelet Coupled model; LST, Land Surface Temperature; NDVI, Normalized Difference Vegetation Index; *T*₃₁, Temperature channel 31; *e*, Emissivity; *P*_{t-clouds}, Mean Top Cloud Cover; *T*_{t-clouds}, Top Cloud Temperature; *H*_{t-clouds}, Mean Top Cloud Height; CO₂, Carbon Dioxide; *R*_v, Radiance; *R*_{clr}(*v*), Clear Sky Radiance; *R*_{bc}, Opaque Radiance; *P*_c, Cloud Pressure; *P*_s, Cloud Surface Pressure; *r*(*v*, *p*), Fractional Transmittance of radiation; *p*, Atmospheric Pressure; O₃, Ozone; SST, Sea Surface Temperature; *σ*_R, Standard Deviation Error; SLFN, Single Feedforward Neural; GCV, Generalized Cross Validation; BOM, Bureau of Meteorology; SILO, Scientific Information for Land Owners; *r*, Correlation Coefficient; CPU, Central Processing Unit; OOB, Out of Bag; RMSE, Root Mean Square Error; MAE, Mean Absolute Error; WI, Willmott's Index; *E*_{LM}, Legates & McCabe's Index; RRMSE, Relative RMSE; RMAE, Relative MAE; RMSE_{ss}, Skill Score based on RMSE; ISR^{for}, Forecasted ISR; ISR^{obs}, Observed ISR; FE, Forecasting Error; MSE, Mean Square Error

* Corresponding authors.

E-mail addresses: ravinesh.deo@usq.edu.au (R.C. Deo), jmi@pku.edu.cn (J. Mi).

also important for renewable energy exploration in data-sparse or remote regions with no established measurement infrastructure but with a rich and viable satellite footprint.

1. Background review

Access to clean and affordable energy remains a critical element of the Sustainable Development Goal 7 of the United Nations Development Program (UNDP 2018–2021). The need to promote renewable and efficient energy technologies that are responsive to population needs in urban and rural areas, for women, men, households and business, is also important for UNDP's goal [1]. Empowered by significant renewable energy potential, Australia's solar energy use was among the top ten nations in 2015. Australian photovoltaic power capacity reached 5.44 GW spread across 1.58 million installations by December 2016 [2], the equivalent of more than one solar panel per person [3]. Due to emerging improved technologies, solar energy extraction costs are reasonably low and large industrial-scale solar power plants provide low-cost electricity compared to fossil fuel and nuclear systems. Australia is expected to reach more than 20 GW of photovoltaic power generation in the next 20 years, equal to one-third of the current total renewable energy generation. This would support the Australian Renewable Energy Target for large-scale renewable electricity generation to reach 33,000 GWh by 2020, and for 23.5% of all electricity to come from renewable resources [4,5].

The highest footprints of incident solar fluxes in Australia exist in desert regions, particularly in the northwest and centre of the continent. Despite a significant push for solar energy utilisation in outlying regions, isolation from Australia's NEM grid is a major challenge. Potential surplus energy that could be harnessed and returned to the existing power grid is an issue of national interest. Considering the high costs of infrastructure required to transfer energy into the existing power grid from remote locations, the design of models that can accurately predict global incident solar radiation (*ISR*) is important for energy policy decision-makers.

To assess long-term solar energy availability (especially over different seasons), a robust predictive model of *ISR* (evaluated at remote or regional sites with high solar potential) is an attractive tool for government, policy-makers, and power investors. It would also support the recent Australian Energy Resource Assessment (AERA) report, which recommends the location of solar power plants be contingent upon the expense of a connection onto the NEM grid. In the intermediate term, the report indicates that future investments will likely focus on isolated solar power grids or nodes that are linked to an existing electricity grid [6]. The dearth of energy predictions for isolated, remote, and regional sites with limited measurement equipment hinders the development of solar energy plants and future minimization of infrastructure costs. Even with the advent of the Bureau of Meteorology's network of high quality ground stations that provide *ISR*, only a pre-selected set of sites are available and related predictors (or model input data such as land surface temperature, cloud cover, or ozone) that can modulate the surface level *ISR* flux are not available at required spatial resolutions.

The use of satellite data to forecast surface *ISR* at regional and continental-scales, and with an inexpensive data approach as adopted in this study, can offer a cost-effective alternative for *ISR* prediction. Recent global studies [7–16] advocate the utility of satellite-variables to predict *ISR*, with most focusing on artificial intelligence methods. In a study utilizing satellite parameters, Linares-Rodriguez *et al.* [12] optimized an artificial neural network (ANN) to validate utility over large spatial areas, applying clear-sky estimates and satellite (Meteosat 9) images as model input variables. Sahin *et al.* [10] developed a fast and simple method to estimate *ISR* in Turkey, tuning an extreme learning machine (ELM) with satellite-based data from the National Oceanic and

Atmospheric Administration Advanced Very High-resolution/radiometer. Evaluated against an ANN model, their ELM generated more accurate estimations of *ISR* for spatially diverse test locations, indicating that the approach could be used to design high-efficiency solar devices. The primary advantage of an approach where several sites with diverse characteristics are used in model training, is feature extraction or mapping which can generate representative, and universally-trained models which can then be applied to any regional site.

Limited input data from Andalusia (Spain) (January–December 2008) incorporated into a genetic algorithm was used to optimize an ANN model, screening eleven 3-km Meteosat 9 channels and one clear-sky term, wherein data from 83 stations across a region were used; 65 for training and 18 independent ones for testing [11]. In another study utilizing statistical methodologies and ANN to estimate *ISR* with NOAA-AVHRR images [12], 661 images of NOAA-AVHRR level 1b were used as model inputs, showing that the ANN model accuracy exceeded the statistical methods. Senkal and Kulai [16] applied ANNs and Meteosat-6 satellite C3 D satellites to estimate *ISR* at 12 different sites in Turkey (26–45°E and 36–42°N), whereas Perez *et al.* [16] used satellite remote-sensing and ground network measurements for the production of site/time specific irradiance with NOAA's intermediate resolution visible channel of geosynchronous satellite. Escobar *et al.* [17] used ground station network and satellite estimations from the Chile-SR model to evaluate its potential for solar photovoltaic power utilisation, and applied a satellite-based model using the Brasil-SR methodology with combined visible and infrared images. Long-term patterns of European PV outputs were modelled with 30 years of validated hourly reanalysis and satellite data, considering MERRA and MERRA-2 global meteorological reanalyses and the Meteosat CM-SAF SARAH satellite in Pfenninger and Staffell [18]. The use of satellite data to improve solar radiation forecasting with a Bayesian ANN model was carried out by Aguiar *et al.* [19], justifying the practical relevance of satellite-based predictors in solar radiation estimation.

Despite a plethora of artificial intelligence models for surface *ISR* estimations throughout the world, application of these predictive models in Australia has been limited, and mechanistic (or physically-based) approaches, that explain physical phenomena, have been used by researchers. Boland *et al.* [20] constructed a mathematical model specifically for Australian weather conditions using the logistic function instead of the previously applied piecewise linear or nonlinear function. Employing a quadratic-programming formulation, a follow-up study by Boland *et al.* [21] designed a statistically rigorous, closed mathematical model for diffuse *ISR* estimation, verifying its usage at two study sites (Adelaide and Geelong). While their model was reasonably accurate for these specific locations a number of challenges were identified, in terms of the significance level for probabilities used to identify outliers, an inability to design a generic model for all locations, and the use of other predictors (e.g., including solar altitude and daily clearness index). To improve previous mathematical models, Huang *et al.* [22] introduced a new *ISR* forecasting approach on an hourly scale by means of a Coupled Auto-Regressive and Dynamical System (CARDS) model, wherein an autoregressive (AR) model with a dynamical system was generated. In this study, the difference between solar radiation at present and at a lag of one time step was used as a correction factor for the predicted values. Applied to one site located in Mildura (Australia), the model was used to simulate 1-h ahead *ISR* forecasts, gaining a 30% improvement in accuracy compared with the model without this correction factor. Despite these studies that provide a promising approach for *ISR* estimation at regional sites major drawbacks remain, particularly in terms of the need for a boundary condition (hence experimental facilities requiring

sufficient data for a mathematical or a statistical approach), limited testing sites where such models have been validated, and the inability of pure mathematical models to address non-linearities in model input variables which would likely influence the model's overall accuracy.

To address the above-identified challenges, moderate resolution imaging spectroradiometer (MODIS) data were used to train an ANN model in a first Australian study by Deo and Sahin [7]. Global solar radiation was predicted in regional Queensland, Australia, where the monthly *ISR* at test sites was modelled with the nearest neighbouring training sites and land surface temperature (LST). Using a limited set of predictor data from 2012 to 2014 for seven groups of sites (three sites per group; for the first two sites, data from 2012 to 2013 were utilized for model development; for the third site, 2014 data was used for cross-validation), the objective model was evaluated against multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) models. ANN outperformed MLR and ARIMA, with 39% of cumulative forecasted errors recorded in the smallest magnitude bracket compared to MLR and ARIMA with 15% and 25%, respectively. In spite of the significant advancement in providing an accurate *ISR* simulation with freely available, remotely-sensed predictors, this study was cross-validated at only a limited set of stations and utilized only the LST (satellite-based predictor) variables.

In another related *ISR* estimation study in Queensland, Deo et al. [23] developed a support vector machine-wavelet coupled model (W-SVM), where ground-based sunshine hours, temperature, wind speed, evaporation, and precipitation were employed as predictor variables. For daily forecast horizons, the wavelet SVM model outperformed the classical SVM model. In spite of the good performance of the MODIS-based ANN for monthly estimation [7] and ground-data based W-SVM model for daily and monthly *ISR* estimation [23], there are two questions that require further consideration: First, *can we design and implement an accurate, data-inexpensive universal learning model where the Australian mainland's continental-scale data are used to evaluate a more representative model;* and second, *can we develop such a learning model with remotely-sensed (rather than ground-based) data, particularly to support solar energy feasibility studies in data-sparse regions?*

The novelty of this paper is the development of a universally-trained extreme learning machine (ELM) model for *ISR* estimation, wherein metropolitan, regional, and remotely-located sites distributed relatively evenly across Australia are employed to train and construct the ELM model. Using a limited set of predictor data (*i.e.*, 2012–2015), the precise aim of this study is as follows: (1) To employ eight carefully screened MODIS Terra radiometer (NASA-built satellite) time-series (*i.e.*, normalized difference vegetation index, land surface temperature, cloud top pressure, cloud top temperature, cloud effective emissivity, cloud top geopotential height, total ozone, and near infrared water vapour) to model surface *ISR* where 26 study sites are used for ELM model design and 15 study sites are used for model cross-validation (testing) purposes; and (2) To evaluate the ELM model's ability to estimate *ISR* at geographically diverse sites; (3) To benchmark the predictive capability of ELM with respect to alternative methods: a random forest (RF), M5 Tree model, and multivariate adaptive regression spline (MARS) model applying diverse statistical score metrics on forecasted and observed *ISR* at test sites. The application of ELM (with universal approximation capabilities) [24] has a distinct advantage over conventional data-driven models (*e.g.* ANN and SVM) in previous studies, due to better accuracy (*e.g.*, [10,25–27]).

This study presents a new way to forecast continental-scale *ISR*; the approach is consistent with previous (yet differently-focused) studies aimed to forecast a target variable in the vicinity of their training sites, distributed spatially across a large region. Currie et al. [28], and Bilgili and Sahin [29], evaluated an ANN model to estimate daily, weekly, and monthly wind speed at a target site nearby the neighbouring sites; Damousis et al. [30] applied a fuzzy model trained with wind speed and direction from neighbouring sites 30 km away from a wind turbine cluster; and Deo et al. [31] predicted target station wind speed using

seven spatially-distributed reference sites with a hybrid multi-layer perceptron integrated with Firefly Algorithm (MLP-FFA) model. The advantage of the Measure-Correlate-Predict strategy, as used previously [32,33], is that target site data can be modelled with features derived from surrounding sites, leading to an increase in the pool of variables required for an accurate estimation but not necessarily requiring a temporally larger dataset. In this respect, satellite data used in the context of the present study can be very useful for regional and data-sparse ground sites where a limited length of data is available.

2. Theoretical background

A description of the satellite data and the relevant theories used to construct the models is presented. Consider the model's training-target data for a study period of length N as $[x \mid y] = [(X_1, X_2 \dots X_k) \mid Y]$ where X (of dimension k) denotes a predictor matrix constructed with remotely-sensed data and Y is the *ISR*. Three years of data were assigned as ELM inputs (2012–2014) denoted with $k = (1, 2\dots, 8)$ indicating the 8 most salient MODIS satellite-based predictors (*i.e.*, land-surface temperature, cloud top pressure, temperature, emissivity, height, ozone concentration, and near-infrared water vapour), 3 site-specific geographic properties (latitude, longitude and elevation, $k = (9, 10, 11)$), and the periodicity representing the monthly cycle ($k = 12$) yielding the $N \times k$ matrix.

The predictor dataset derived from the MODIS Terra satellite sensors is described below in Section 2.1.

2.1. Moderate resolution imaging spectroradiometer satellite product

To construct a universally-deployable, continental-scale model with remotely-sensed predictors, monthly inputs captured by the MODIS Terra radiometer (a NASA-built satellite) are extracted. Satellites are conveniently used to model ground-based targets with remote-sensing techniques [7,10,34–38]. The earth's radiative properties are quantified through an atmospheric window without establishing a physical connection with the ground-based target variable [39]. MODIS data used to model a target variable provides an added opportunity to predict ground events without the need for infrastructure or capital costs such as expensive measurement instrumentation for regular monitoring, particularly for large spatial-scale phenomenon in remote and regional locations where measurement facilities do not exist.

Currently, to acquire remotely-sensed data [40], several satellites are used that include the Geosynchronous Meteorological, Meteosat, Insat, Goes, National Oceanic and Atmospheric Administration (NOAA) series, Advanced Spaceborne Thermal Emission Reflection Radiometer, Fengyung-1C, D, and Metop. However, the application of data from each satellite may differ based on the problem of interest and spatial positioning of the satellite where the application of that data is required. To convert satellite records to an interpretable quantity for model development, data conversion algorithms for the Advanced Very High Resolution Radiometer (AVHRR) sensors have been developed. These algorithms are applied to extract remotely-sensed data for different model development [40–44].

In this study, the authors utilize the MODIS Terra (EOS AM) (launched 18 December 1999) as this satellite has an identifiable footprint in Australia [7]. It passes from north to south across the equator every morning (approximately 10.30 A.M.) and views the entire earth's surface every 1–2 days. Therefore, MODIS EOS AM is able to acquire data in a total of 36 different spectral bands, or a group of wavelengths, with $\pm 55^\circ$ scanning patterns at an orbit of 705 km and with about a 2330 km swath width [45–47]. The sun-synchronous orbit of this satellite allows it to pass over the same area at the same time in every 24-h period (at every 99th minute orbit the satellite crosses the equator).

To model ground-based *ISR*, the eight most salient variables are obtained from the MODIS Terra (EOS AM) satellite. A brief description

of each variable and the respective algorithm to convert the image into a quantifiable variable is described. Included in the MOD13C2 data product [48] is the normalized difference vegetation index (*NDVI*), representing a consistent, spatial, and temporal comparison of the global vegetation conditions that can be used for monitoring earth's terrestrial photosynthetic vegetation activity, change detection, biophysical interpretation, and an indication of the solar radiation. The compositing algorithm for *NDVI* utilizes a pixel-based bidirectional reflectance distribution function to normalize the reflectance to a nadir view and a standard solar angular geometry. *NDVI* is a 'normalized' transformation of the non-infrared (*NIR*) to red reflectance ratio, $\rho_{\text{nir}}/\rho_{\text{red}}$, designed to standardize vegetation indices to be between -1 and +1 [48]:

$$\text{NDVI} = \frac{[\rho_{\text{nir}}/\rho_{\text{red}} - 1]}{[\rho_{\text{nir}}/\rho_{\text{red}} + 1]} \quad (1)$$

where ρ is the reflectance ratio.

Monthly land surface temperature (*LST*), a satellite parameter closely associated with *ISR* [7], is available from the MODIS Terra sensor's 31st and 32nd channels. The two channels have a spectral radiative band of 10.78–11.28 and 11.77–12.27 μm , respectively [45,46]. The MODIS/Terra Land Surface Temperature and Emissivity (*LST/e*) products provide per-pixel temperature and emissivity values in a sequence of swath-based to grid-based global products. MODIS Terra *LST/E* Monthly L3 Global 0.05Deg CMG (Short name: MOD11C3) used in this research, is configured on a 0.05° latitude/longitude climate modelling grid (CMG). To acquire an accurate estimate of the *LST*, the atmospheric effect can be minimized by a well-known split-window formula as per the study of [49], and used in global solar energy modelling [7,35].

An error threshold lower than 1 K is set for *LST* calculations bounded by -10 K and -50 K using the approach of Wan and Dozier [49]. Less than 1 K error is usually evident in the homogenous land surface terrains, verified in earlier studies [7,50–52]. The algorithm [49] for this study is:

$$\begin{aligned} \text{LST} = & \left[A_1 + A_2 \frac{1 - \varepsilon}{\varepsilon} + A_3 \frac{\Delta\varepsilon}{\varepsilon^2} \right] \frac{T_{31} + T_{32}}{2} \\ & + \left[B_1 + B_2 \frac{1 - \varepsilon}{\varepsilon} + B_3 \frac{\Delta\varepsilon}{\varepsilon^2} \right] [T_{31} - T_{32}] + C \end{aligned} \quad (2)$$

where T_{31} and T_{32} denote the brightness temperature recorded by the MODIS Terra Channel 31 and 32, respectively, and A_i & B_i ($i = 1, 2, 3$), and C are the constants depending on the point of viewing angle between 0° and 65°, which is dependent on air-ground temperature and water vapour captured by the MODIS sensor with regression formulas, and ε is the emissivity, while $\Delta\varepsilon$ is the difference of the emissivity, written as:

$$\varepsilon = 0.5(\varepsilon_{31} + \varepsilon_{32}) \quad (3)$$

$$\Delta\varepsilon = \varepsilon_{31} - \varepsilon_{32} \quad (4)$$

where ε_{31} and ε_{32} are the emissivities of the 31 and 32 channels [46,49].

Algorithms used for the detection and quantification of mean cloud top pressure, temperature, and height ($P_{t\text{-clouds}}$, $T_{t\text{-clouds}}$ and $H_{t\text{-clouds}}$) are available in MODIS Cloud Product (MOD08M3), as described by Menzel et al. [53]. This approach utilizes both infrared and visible data reduction techniques to determine the physical and radiative cloud properties [54], allowing a detection of the cloud-particle phase (ice vs. water, clouds vs. snow), effective cloud-particle radius, and cloud optical thickness derived from MODIS visible and near-infrared channel radiances. They also cater to cloud shadows affecting a viewed spot. In addition to the $P_{t\text{-clouds}}$ data, this algorithm also generated satellite data for $H_{t\text{-clouds}}$, $T_{t\text{-clouds}}$, and $\varepsilon_{t\text{-clouds}}$, including the ice vs. water, opaque vs. non-opaque, and cloud fraction using the infrared retrieval methods both day and night at 5 × 5 grid 1-km-pixel resolution [54–56].

Cloud top properties such as cloud height, temperature, and

effective emissivity, which is intrinsically related to *ISR*, are generated using a CO₂ slicing algorithm. This algorithm aims to correct for possible cloud semi-transparency effects [53]. In this regard, the MODIS-based infrared CO₂ channel can be used to track the clouds at a 5 × 5 pixel resolution, generating a global cloud climatology at approximately 1.0° resolution. MODIS bands have a spatial resolution of 1 km at nadir and an *IR* cloud phase is inferred from the MODIS 8.5 and 11 μm brightness temperatures at 5 × 5 pixel resolution, and then temporally averaged at 1.0° resolution. A CO₂ slicing algorithm is used for cloud detection and quantification based on the atmosphere that becomes more opaque due to CO₂ absorption as the wavelength increases from approximately 13.3 to 15 μm .

The CO₂ slicing technique adopts a radiative transfer process in an atmosphere with a single cloud layer. For a given cloud element in a field of view, the radiance (R_o) observed is written as [53]:

$$R(v) = (1 - N \times E) R_{\text{clr}}(v) + N \times E \times R_{\text{bcd}}(v, P_c) \quad (5)$$

where $R_{\text{clr}}(v)$ is the clear sky irradiance, $R_{\text{bcd}}(v, P_c)$ is the opaque irradiance from pressure level P_c , N is the fraction of the field of view covered with cloud, and E is the cloud emissivity. Note that R_o is computed as a sum of two terms since MODIS sensor results are separated by cloud type (*i.e.*, clear sky with infrared window optical depth $\delta < 0.05$, thin clouds with $\delta < 0.70$, thick clouds with $\delta > 0.7$, and opaque at $\delta > 3.0$ [53,57]). Accordingly, the total effect of clouds (by a sum of two terms in Eq. (5)) is taken into account by CO₂ slicing techniques for a given cloud element in the field of view.

The opaque cloud irradiance can be deduced as [53]:

$$R_{\text{bcd}}(v, P_c) = R_{\text{clr}}(v) - \int_{P_c}^{P_s} \tau(v, p) \frac{dB[v, T(p)]}{dp} dp \quad (6)$$

where P_s is the surface pressure, P_c is the cloud pressure, $\tau(v, p)$ is the fractional transmittance of radiation of frequency v emitted from the atmospheric pressure level (p) arriving at the top of the atmosphere ($p = 0$), and $B[v, T(p)]$ is the Planck radiance of frequency v for temperature $T(p)$. $P_{t\text{-clouds}}$ is derived from radiance ratios between two spectral bands that apply the ideas of Chahine [58] and Smith and Platt [59]. The ratio of the deviations in the observed radiances, $R(v)$ to their corresponding clear-sky radiances $R_{\text{clr}}(v)$, for two spectral bands of frequency v_1 and v_2 in the same field of view, is written as [53]:

$$\frac{R(v_1) - R_{\text{clr}}(v_1)}{R(v_2) - R_{\text{clr}}(v_2)} = \frac{NE_1 \int_{P_s}^{P_c} \tau(v_1, p) \frac{dB[v_1, T(p)]}{dp} dp}{NE_2 \int_{P_s}^{P_c} \tau(v_2, p) \frac{dB[v_2, T(p)]}{dp} dp}. \quad (7)$$

In Eq. (7), for those frequencies that are spaced closely in wavenumber, the approximation $E_1 \approx E_2$ is assumed, thus allowing the pressure of the cloud within the field of view to be specified.

Using the $P_{t\text{-clouds}}$ data, an effective cloud amount, also called the effective emissivity, $\varepsilon_{t\text{-clouds}}$, can be evaluated from the infrared window band where the following relationship is applied [53]:

$$NE = \frac{R(w) - R_{\text{clr}}(w)}{B[w, T(P_c)] - R_{\text{clr}}(w)} \quad (8)$$

In Eqs. (7) and (8), N represents the fractional cloud cover within the field of view, NE is the effective cloud amount, w is the window band frequency, and $B[w, T(P_c)]$ is the opaque cloud radiance.

It is important to note that the effective cloud amount is restricted to the accurate estimation of the window band of the clear sky radiance. Therefore, when $NE < 1$, the satellite sensor could be observing a broken cloud $N < 1$, $E = 1$, overcast transmissive cloud $N = 1$, $E < 1$, or broken transmissive cloud $N < 1$, $E < 1$ [60].

The present study has utilized MOD08M3, as this particular repository has distinct advantages such as the inclusion of cirrus reflectance in the visible spectrum at a 1-km-pixel resolution, used to remove scattering effects. This can enhance the accuracy of the estimated atmospheric properties. The relevant cloud top properties (*e.g.*, $P_{t\text{-cloud}}$) can be obtained from the MODIS Terra Channels 28, 29, and

31–36. The algorithm also applies a cloud-mask (MOD35) that acts as a screen for cloud data, and the global model analyses of sea surface temperature (SST) and pressure, including the profiles of temperature and moisture and the blended SST, are applied to calculate cloud forcing with respect to atmospheric pressure and emissivity. The infrared cloud phase algorithm is applied to Channels 28, 29, and 31 and a validation of cloud top heights are conducted through comparisons with CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) data and other Lidar estimates using aircraft observations [55,56,61]. Cloud emissivities ($\epsilon_{\text{t-clouds}}$) are available as monthly composited averages [62,63]. For cross-validation, these data are compared with Lidar-determined cloud optical thickness, used in similar studies utilizing remotely-sensed data (e.g., [55,56,61]).

ISR is also related to the total ozone (O_3) and near infrared water vapour dynamics (H_2O) available in the MOD08M3 product, obtained as a monthly composited average [64]. The algorithm for the retrieval of O_3 , including vertical profiles (soundings) of temperature and moisture, integrated total column precipitable water vapour, and several other atmospheric stability indices, are outlined by Seemann et al. [65] and Borbas et al. [66]. The O_3 data are available in MODIS Terra Channel 30 (band width 9.580–9.880 μm) whereas the near infrared water vapour is available in MODIS Terra Channel 27–29 (band width 6.35–8.70 μm). For details, readers can consult the current literature (e.g., [55,64]).

2.2. Objective model: extreme learning machine

This study integrates satellite products into an extreme learning machine (ELM), an AI algorithm with a state-of-the-art single layer feed forward network (SLFN) for the input of satellite data and feature optimization (or learning space). The ELM model performs regression analysis with respect to the measured *ISR* and an output space where the forecasted *ISR* is generated [67]. As an improved algorithm originating from ANNs, the generalized ELM model employs least squares estimations relying on the weights and biases attained in the learning space. This random assignment of hidden neurons enables the ELM algorithm to solve regression problems much more quickly and with greater accuracy than the other learning methods (e.g., ANN, SVM) [68]. Assigning the most relevant weights (and biases) in a random fashion with outputs based on least-squares, the ELM utilizes the Moore-Penrose inverse function to forecast *ISR*, thus avoiding iterative training (as with ANN or SVM) and yielding solutions that collapse to a global, rather than local, minimum [24]. Advantages of ELM include a faster convergence rate, superior generalization, lack of local minima issues, less over-fitting, and no iterative tuning. In accordance with universal approximation theory randomly initiated neurons remain fixed, hence the model is remarkably efficient [24,67,68].

A three-stage process is evoked: (I) generation of hidden layer weights and biases, (II) inputs that are passed through the hidden layer parameters to create a hidden layer output matrix, and (III) the hidden layer inverted to the target matrix (viz., Moore-Penrose generalized inverse) where a system of linear equations is solved, utilizing the interconnections between satellite-based predictor features and the measured *ISR* data. Following earlier studies [24,67,68], for $i = 1, 2, \dots, N$, the SLFN with L hidden neurons is expressed as:

$$\psi_L(\mathbf{x}) = \sum_{i=1}^{i=L} h_i(\mathbf{x}) \cdot \beta_i = h(\mathbf{x})\beta \quad (9)$$

where,

β i.e., $[\beta_1, \beta_2, \dots, \beta_L]^T$ is the target weight matrix for the hidden and output neurons.

$h(\mathbf{x})$ i.e., $[h_1, h_2, \dots, h_L]$, is the hidden neuron outputs of satellite-derived predictors \mathbf{X}_i ,

$h_i(\mathbf{x})$ is the i th hidden neuron, which can be written as

$$h_i(\mathbf{x}) = \vartheta(\mathbf{a}_i, b_i, \mathbf{X}) \text{ and } \mathbf{a}_i \in R^d, \quad b_i \in R \quad (10)$$

In learning space, non-linear, piecewise-continuous hidden layer activation function $h_i(\mathbf{x})$ must be defined for neuronal parameters (\mathbf{a}, b) that satisfy the universal approximation theorem, $\vartheta(\mathbf{a}_i, b_i, \mathbf{X})$.

$$\text{Sine } \vartheta(\mathbf{a}, b, \mathbf{X}) = \sin(\mathbf{a}\mathbf{X} + b) \quad (11.1)$$

$$\text{Tangent Sigmoid } \vartheta(\mathbf{a}, b, \mathbf{X}) = \frac{2}{1+e^{-2(-\mathbf{a}\mathbf{X}+b)}} - 1 \quad (11.2)$$

$$\text{Logarithmic Sigmoid } \vartheta(\mathbf{a}_i, b_i, \mathbf{X}) = \frac{1}{1+e^{(-\mathbf{a}\mathbf{X}+b)}} \quad (11.3)$$

$$\text{Hard Limit } \vartheta(\mathbf{a}, b, \mathbf{X}) = 1 \text{ if } \mathbf{a}\mathbf{X} + b > 0 \quad (11.4)$$

$$\text{Triangular Basis } \vartheta(\mathbf{a}, b, \mathbf{X}) = 1 - |\mathbf{a}\mathbf{X} + b| \text{ if } 1 \leq \mathbf{a}\mathbf{X} + b \leq 1, \text{ or } 0 \quad \text{otherwise} \quad (11.5)$$

$$\text{Radial Basis } \vartheta(\mathbf{a}, b, \mathbf{X}) = e^{[-(-\mathbf{a}\mathbf{X}+b)^2]} \quad (11.6)$$

The trained model's approximation error must be minimized when solving for the weights connecting hidden and output layers (β) with a least squares method:

$$\min_{\beta \in R^{L \times m}} \|H\beta - T\|^2 \quad (12)$$

where,

$\|\cdot\|$ is the Frobenius norm, and

H is the hidden layer output matrix, given as:

$$H = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_N) \end{bmatrix} = \begin{bmatrix} g_1(a_1x_1 + b_1) & \dots & g_L(a_Lx_1 + b_L) \\ \vdots & \dots & \vdots \\ g_1(a_Nx_N + b_1) & \dots & g_L(a_Nx_N + b_L) \end{bmatrix} \quad (13)$$

T is the target matrix, drawn from the training dataset, and given as:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & & \vdots \\ t_{N1} & \dots & t_{Nm} \end{bmatrix} \quad (14)$$

An optimal solution is then determined by solving a system of linear equations:

$$\beta^* = H^+ T \quad (15)$$

where H^+ is the Moore-Penrose generalized inverse function (+).

2.3. Benchmark model 1: Random forest

Random forest (RF), adopted to benchmark the ELM, is a method that relies on model aggregation [69]. RF is a family of ensemble data mining tools that avoid overfitting of the trained model. To facilitate *ISR* forecasts with satellite-based predictors, the RF algorithm integrates binary decision trees built with bootstrapped samples acquired from a learning sample D (i.e., predictor-target matrix, $[\mathbf{x} | \mathbf{y}]$ where a subset of explanatory variable \mathbf{X} has been screened randomly at each node). In an ensemble-based RF algorithm, a combination of trees (typically up to 2000) are grown based on the predictor dataset wherein each tree is generated from a bootstrap sample while leaving one third of the samples for validation or the “out-of-bag” (OOB) forecasts.

This process involves branching a set of decision trees drawn on a randomized subset of satellite-based predictors, and the final outcome is determined as a mean of all decision trees [69,70]. RF employs OOB samples to determine trained model error, associated with an independent observation set utilized to evolve the decision tree. No separate cross-validation data (as with the case of an ANN model) are required [71].

The RF algorithm can be explained in four stages:

- a. Utilizing the predictor-target matrices $[x | y]$ with N values in the training set, a bootstrapping procedure is applied, drawing randomly with replacement used for the evolution of original decision trees.
- b. Given k predictors that are believed to govern the behaviour of *ISR*, the RF algorithm will specify a number $K < k$ such that at each node of the decision tree, K variables are drawn randomly out of the k dataset. The best split on the K variables is adopted while holding K as a constant as the randomized forest evolves.
- c. During the RF execution process, each decision tree is first grown to its maximum extent without a pruning process applied to reduce overall structure.
- d. The unseen (*i.e.*, *ISR*) data are predicted by aggregating the K tree forecasts (*i.e.*, the average value of all trees is determined to generate the forecasted *ISR* values in an independent test set).

When modelling a given dataset, the OOB error estimates can be accurate as long as a sufficient number of decision trees are grown to enable multi-model aggregation [72].

2.4. Benchmark model 2: M5 Tree Model

M5 Model Tree [73], adopted to compare ELMs, is a hierarchical model based on a binary decision framework. The model employs a linear regression approach applied at the terminal (leaf) nodes to establish the cause and effect relationships in the predictor-predictand matrix [74]. To construct an M5 Tree, a dual process is applied where the $[x | y]$ training data pairs are split into subsets to create decision-trees for pattern recognition [75]. This is facilitated recursively in such a way that the subsets are created following a criterion which depends on the standard deviation of class values, while noting the reduction in model error, σ_R [76,77]:

$$\sigma_R = \sigma(\Lambda) - \sum_{i=1}^{i=N} \frac{\Lambda_i}{\Lambda} \sigma(\Lambda_i) \quad (16)$$

where:

Λ is a set of data samples that escalate to the node,
 Λ_i is the subset of data samples that have the i th outcome of a potential test set.

Tree splitting continues to provide data attribute extraction until the optimal number of splits are attained. A regression is applied to screen the tree rules for an optimal M5 model (*i.e.*, a model that yields the minimum error, σ_R). If a model is constructed from several data points with a multiple predictor-target matrix, smoothing must be applied to compensate for abrupt discontinuities between adjacent models at the leaves of a pruned tree [76,77], thereby improving model accuracy. Quinlan [73] and Witten and Frank [78] provide further details of the M5Tree model.

2.5. Benchmark model 3: MARS

Multivariate adaptive regression spline (MARS), a non-neural based method [79], is also adopted to benchmark ELM. MARS considers multivariate matrices (*i.e.*, satellite predictors *vs ISR*) to analyse predictor contribution measured with basis functions, where interactive effects of explanatory behaviour are used in forecasting *ISR* [80]. MARS determines the importance of each individual (and subsequently added) predictor, allowing for complex and non-linear relationships to be identified and the best regressors to be selected [81,82]. A significant merit of MARS is that it makes no assumptions about the relationships between predictors and *ISR* [79,83]. Instead, it generates forecasts based on learned relationships [79,84]. Data are characterized globally by cubic or linear equations, providing the MARS model the flexibility

to consider bends, thresholds, and departures from linearity [85,86]. Splines ensure adequate data in a subgroup, therefore avoiding overfitting by using minimum distance between neighbouring knots. The basis functions, $BF(x)$, are determined from satellite-based predictors and projected on the *ISR* [82,86].

If \mathbf{X} is a k -dimensional vector, the MARS model is denoted as:

$$\mathbf{Y} = f(\mathbf{X}) + \xi \quad (17)$$

where:

N is the number of training datum points, and
 ξ is the distribution of error [77,81].

MARS approximates $f(\mathbf{X})$ by applying $BF(x)$ with a piecewise function: $\max(0, x - c)$ and c represents the knot's position [85].

The term $\max(.)$ indicates that only the positive part is used; otherwise it is zero. This is written as:

$$\max(0, x - c) = \begin{cases} x - c & \text{if } c \geq t \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

$f(\mathbf{X})$ is then constructed as a linear combination of $BF(x)$

$$f(\mathbf{X}) = \beta_0 + \sum_{n=1}^{n=N} \beta_n BF(\mathbf{X}) \quad (19)$$

where β = constant estimated using least-squares and $f(\mathbf{X})$ is applied as a forward-backward method to identify knots [82].

The backward phase uses the Generalized Cross Validation (GCV) statistic (*i.e.*, a standardized measure of mean square error, MSE) for the deletion of the redundant $BF(x)$ to prune the model [87–89]:

$$GCV = \frac{MSE}{\left(1 - \frac{enp}{N}\right)^2} \quad (20)$$

where,

enp = effective number of model parameters, $enp = k + c (k - 1)/2$
 k = basis functions in the MARS model (incl. intercept term)
 c = penalty (set to 2 or 3)
 $(k - 1)/2$ = hinge-function knots

When $enp \geq N$, GCV is set to *Inf* (*i.e.*, the model is infinitely bad) [90].

3. Materials and methods

3.1. Context and importance of the study sites

This paper constructs and evaluates ELM *vs.* RF, M5 Tree, and MARS models for large spatial-scale forecasting of global incident solar radiation (*ISR*), integrated with satellite-derived MODIS Terra predictor data [45] applied to a set of regularly distributed sites (including regional areas and metropolitan cities) throughout Australia. A large spatial-scale solar prediction model is important for this nation since Australia has one of the world's highest rates of solar radiation per square metre, and annually receives about 58 million petajoules, 10,000 times the country's annual energy consumption. Large-scale solar electricity is rapidly expanding with more than 2 million households using a roof-top based system [6,27]. Solar potential is high in north-west and central Australia, areas that do not have access to the national electricity grid. The evaluation of models in data-sparse sites can enable strategic decision-making for future locations of solar powered sites, utility-scale plants, and transmission infrastructures for energy networks. In accordance with the Australian Energy Resource Assessment report [6,91,92], photovoltaic systems powered primarily by received *ISR* are well suited to off-grid electricity generation and can

complement the energy supply at sites where electricity generation costs are high (particularly for remote and regional communities). In the Pilbara region in Western Australia, for example, the installation of photovoltaic power systems is approaching parity with diesel powered systems for off-grid energy supply.

Fig. 1 shows the annual solar energy potential across Australia, as measured by the *ISR* ($\text{MJ m}^{-2} \text{ day}^{-1}$) using January–December 2016 data from the Australia Water Availability Project [93,94]. It can be seen that from September to February, the mean annual solar radiation is more than $25 \text{ MJ m}^{-2} \text{ day}^{-1}$.

Dominant regions with high *ISR* include the north-west deserts and the centre of Australia (**Fig. 1**). These correspond to areas without a well-established National Electricity Grid. Wyld Group and MMA [95] have identified several regional locations suitable for solar power plants based on high solar radiation fluxes and proximity to local loads. These regions have a high electricity cost. Within the National Electricity Market grid area, Port Augusta (South Australia), north-west Victoria, and central and north-west New South Wales (including Mount Isa, Alice Springs and Tennant Creek), have been identified as having high potential for solar thermal powered investments, which also corresponded to their high levels of *ISR* (**Fig. 1**). Kalbarri in the vicinity of Geraldton (Western Australia), the South-West Interconnected System, the Darwin-Katherine Interconnected System, and Alice Springs–Tennant Creek can also be considered potential solar powered sites for evaluating solar energy assessment models and assessing the long-term feasibility of developing solar power in these areas.

Considering the need to expand solar energy networks, not only for capital or metropolitan cities where they are currently largely concentrated, but also in sub-metropolitan areas and more importantly in

regional and remote sites identified as ‘very solar-rich’, the evaluation of *ISR* forecast models over spatially sparse sites (e.g., desert and central) is a significant problem of interest for future solar energy investment and policy development.

3.2. Data and modelling approach

To develop an ELM model using a limited set of satellite predictor datasets for *ISR* prediction, that is also tailored to all of Australia, this study first screened approximately uniformly distributed study sites with reliably measured surface level *ISR* scattered over 41 locations. **Fig. 2** plots the spatial distribution of the sites and shows a total of 41 training and testing locations. Since one primary purpose is to forecast long-term *ISR* using a limited set of satellite-derived data, predictor data from four years (2012–2015) are extracted from the MODIS Terra sensor of NASA-satellites. *ISR* are obtained for the surface level observation towers from Scientific Information for Land Owners (SILO) [27,45,96].

The Australian Bureau of Meteorology (BOM) has made *ISR* data publicly available and regularly assesses the reliability of their solar energy networks. This ensures that an accurate, effective, and cost-efficient mechanism is in place for *ISR* data acquisition [97]. The SILO database is developed by the Queensland Department of Environment and Resource Management from BOM-provided observational *ISR* records, where missing values have been inputted in accordance with statistical procedures [98–100]. This has enabled SILO-based *ISR* data to be utilized in previous solar energy-related studies [7,23,27].

Considering the need to forecast *ISR* precisely with a largely universal, spatially-relevant model validated over both regional and

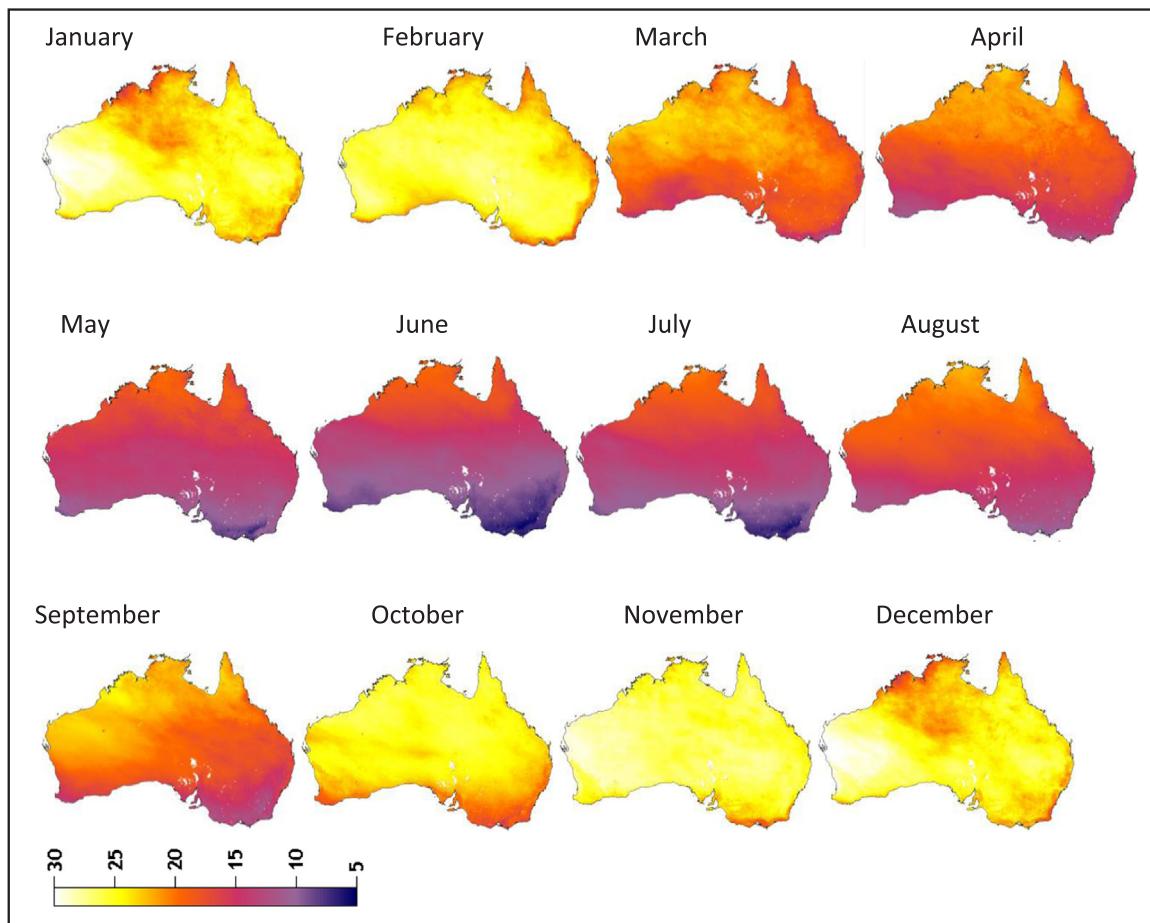
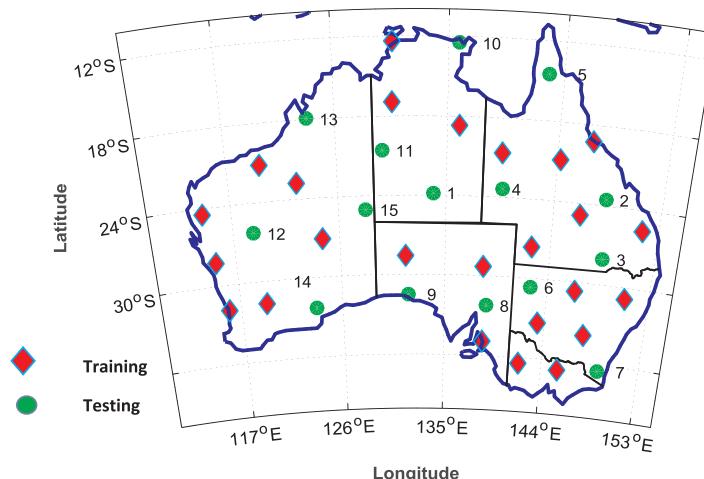


Fig. 1. Annual solar energy potential measured in terms of Incoming Solar Irradiance, *ISR* ($\text{MJ m}^{-2} \text{ day}^{-1}$) over Australia using data from January to December 2016, from the Australia Water Availability Project [1,2].



meteorological locations, this study partitioned selected data from 41 locations into a subset of 26 study sites for model development, and the remainder (15 study sites) for model cross-validation (*i.e.*, testing). The testing locations are spread uniformly across Australia in such a way that a universal predictive model based on any learning algorithm (*i.e.*, ELM, RF, MT Tree, or MARS) can be constructed with data from the 26 sites pooled into a single predictor-target matrix.

The universally-trained and cross-validated model is a novel contribution in the context of this prediction problem, expanding the methodology of the author's previous study [7]. The previous study utilized an ANN algorithm to predict *ISR* for regional study sites in Queensland. However, that study incorporated 7 groups (*i.e.*, 3 station members each) as training and cross-validation sites where data for 2 stations had been used to train a regionally-specific universal model and data for 1 station had been used to cross-validate (*i.e.*, test) the universal model. The present study, however, adopts a large-scale, spatially-constructed universal model with MODIS Terra satellite inputs for 41 Australia-wide sites (for effective feature extraction). This generates an optimal predictive tool that is cross-validated for the remaining 15 testing sites.

In Table 1, geographic properties of the training and cross-validation sites are enumerated, including mean annual statistics of MODIS Terra satellite-derived and *ISR* data from 2012 to 2015. To confirm agreement between predictors and target variable, Fig. 3 displays a scatterplot of monthly MODIS-based predictors and solar radiation at training sites. As with earlier studies (*e.g.*, [7]), land surface temperature pooled for Australia attained the highest coefficient of determination ($r^2 = 0.7924$) with respect to the target (*ISR*). The other significant predictors are found to be: cloud emissivity ($r^2 = 0.1679$), normalized vegetation difference index, ($r^2 = 0.1526$) and clear-sky near infrared water vapour ($r^2 = 0.1376$). The role of cloud top height and cloud top mean pressure are found to be relatively small (*i.e.*, with only 27.5% and 30.4% of covariance in the *ISR* explained by these predictors), while the role of cloud top temperature is the weakest with an r^2 of 0.00106 (Fig. 3).

Fig. 4 is a correlogram constructed between *ISR* and satellite data pairs in the training phase; it demonstrates a relationship between paired variables in terms of cross-correlation coefficient r_{cross} . The blue lines indicate statistical significance of r_{cross} at a 95% confidence interval with respect to the covariance at different lagged values. The successively time-shifted data of predictor variables with respect to target *ISR* reveal a high degree of serial correlation ($-0.25 \leq r_{\text{cross}} \leq 0.95$) at the different monthly antecedent values, except for cloud top temperature where the magnitude of r_{cross} is small and marginally significant. Although cloud top temperature is not highly correlated at a lag of $t = 0$ and 2 months, there appears to be a statistically significant

Fig. 2. Spatial distribution of 41 study sites (26 sites in red for model construction and 15 sites in green for model testing) adopted in forecasting *ISR* over the Australian continent. Note: training sites used 3 years of data (2012–2014) while testing sites used a single year (2015). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Test Study Sites

1. Alice Springs
2. Emerald Radar
3. St George
4. Marion Downs
5. Strathburn
6. White Cliffs
7. Nimmitabel
8. Belton
9. Yatla
10. Lake Ewella
11. Tanami Downs
12. Munarra
13. Udailla
14. Noondoonia
15. Giles Office

cross-correlation between lags of 3–8 months, and hence, it is likely to contain predictive features that enable forecasting of global incident solar radiation. The pool of eight satellite-based predictor variables (*i.e.*, Fig. 4) derived from the MODIS-satellite data product is considered to be suitable in terms of exploratory variables providing input data attributes necessary for developing a universal *ISR* prediction model.

The significance of periodicity in climate variable predictions is well-known [7,101], as monthly cycles are a pertinent indicator of seasonality that can enhance the predictive model. Fig. 5 illustrates how cyclic patterns in MODIS Terra satellite-based predictors are related to the objective variable. Training data from all study sites (Table 1) are pooled into a global matrix averaged over the entire study period (2012–2015) and categorized in different months to depict seasonality. It is clear that *LST* and *NDVI* are very closely (albeit inversely) correlated to *ISR* in terms of monthly variations, wherein the highest and lowest values for *LST* and *NDVI* are noted in December–January and June–July, respectively. The cloud top mean pressure is inversely correlated with *ISR* whereas the near infrared water vapour and mean height of top clouds follows a similar pattern to *ISR* data. The lowest values noted for the latter two variables are in the period of July and August although the lowest values for *ISR* data are found in June, averaged for the three-year period.

The extent of agreement between cloud effective emissivity and ozone concentration with surface level *ISR* is lower than the other variables, and the time-lagged correlation (Fig. 4) does not reveal statistically significant relationships (especially between 3 and 7 months and the 2–5 months lagged data). Considering the reasonable degree of association between MODIS Terra satellite-based predictors and *ISR*, the eight selected variables are considered appropriate predictors to develop a universally-trained ELM model over spatially represented sites in Australia.

3.3. Model development

ELM and the comparative models (*i.e.*, RF, M5 Tree, and MARS) are developed in a Windows 10 platform using MATLAB 2017b sub-routines running on a PC equipped with an Intel(R) core i7–4770 CPU 3.4 GHz. In order to predict *ISR* at 15 test sites across Australia, the ELM model draws on patterns embedded within the k ($= 8$) lots of satellite-based predictor data (including 3 geographic and 1 periodicity-based) input and their relationships with objective variables (Table 1). The performance is benchmarked against RF, M5 Tree, and the MARS model. Fig. 6 shows an ELM model schematic.

The model is constructed with a three-layer neuronal arrangement with an input-hidden-output node system (Fig. 6(a)) [27,102,103]. The input layer contains a satellite and geo-temporal dataset while the

Table 1
Geographic and climatic statistics of the 41 training and testing study sites used to derive satellite-based predictor variables and global incident solar radiation (ISR) over the period 2012–2015. Sources of predictors and target data are as follows: Geographic Inputs and ISR [96]; MODIS Terra satellite [45].

Site #	Name BOM ID	Geographic Inputs		MODIS Terra Satellite Input(s)						Mean Annual Statistics	Target	
		Location	Elevation (m)	NDVI	Land Surface Temperature (LST), (K)	Top Cloud Pressure (P_{cloud}) (Pa)	Top Cloud Temperature (T_{cloud}) (K)	Cloud Effective Emissivity (ϵ_{cloud})	Top Cloud Geopotential Height (H_{cloud}) (hPa)	Total Ozone (O_3)	Near Infrared Water Vapour (H_2O)	
26 Training Sites												
1	Adelaide, 23034	34.95°S; 138.52°E	2	0.40	296.78	702.41	233.92	0.63	3942.00	331.04	1.82	17.83
2	Attunga 055000	30.91°S; 150.86°E	369	0.46	300.35	697.50	236.34	0.58	3995.92	317.70	1.81	19.03
3	Ballarat, 89002	37.51°S; 143.79°E	435	0.36	306.45	760.58	253.15	0.57	3506.67	276.97	3.18	21.06
4	Brunette, 15085	18.64°S; 135.95°E	218	0.48	293.53	698.44	229.17	0.71	3955.33	346.06	1.68	15.54
5	Byrock, 048021	30.66°S; 146.41°E	152	0.46	304.85	663.30	236.70	0.57	3955.33	299.69	2.41	19.33
6	Charles Point, 14031	12.39°S; 130.63°E	14	0.21	312.60	673.75	241.74	0.50	4392.44	292.03	2.49	20.97
7	Euroa, 82016	36.40°S; 141.88°E	140	0.25	314.10	683.84	246.48	0.50	4288.77	292.03	2.53	22.34
8	Gayndah, 39323	37.51°S; 143.79°E	435	0.52	312.96	695.58	227.54	0.58	4179.90	275.62	2.60	22.27
9	Gerang, 78013	36.40°S; 141.88°E	140	0.17	310.46	712.00	243.70	0.58	4179.90	300.80	2.14	21.18
10	Great Vic., 16102	28.53°S; 131.57°E	295	0.41	304.24	638.90	227.90	0.51	4902.04	301.51	1.93	19.47
11	Hughenden, 30022	20.82°S; 144.23°E	316	0.51	301.25	636.76	229.16	0.61	4760.56	322.03	1.58	18.74
12	Jackson Air, 045068	27.58°S; 142.38°E	193	0.30	304.74	685.88	233.72	0.61	4237.98	334.17	1.58	18.81
13	Marble Bar, 4106	21.18°S; 119.75°E	1	0.44	301.33	645.97	226.34	0.61	4709.21	333.92	1.71	17.81
14	Mardathuna, 6032	24.47°S; 114.56°E	90	0.56	298.42	637.53	223.54	0.67	4646.06	340.57	1.30	16.67
15	Marree, 17076	29.28°S; 138.38°E	43	0.37	299.30	659.57	226.46	0.60	4503.65	342.61	1.71	16.83
16	Mount Isa, 029127	20.68°S; 139.49°E	340	0.16	309.30	689.68	238.04	0.51	4273.60	301.90	2.11	20.95
17	Mt Sanford, 14865	16.98°S; 130.56°E	220	0.20	310.97	642.26	230.48	0.58	4860.96	291.39	1.99	20.43
18	Parmgur, 13043	22.82°S; 143.40°E	400	0.25	314.53	670.46	243.32	0.59	4634.79	266.04	2.74	22.56
19	Perth, 09021	31.93°S; 115.98°E	15	0.33	313.66	610.86	230.49	0.51	5616.90	258.39	3.09	22.36
20	Pooncarie, 047024	33.34°S; 147.62°E	75	0.51	304.58	548.29	209.90	0.50	7123.17	261.13	3.96	21.35
21	Quandialla, 73145	143.40°E 142.66°E	225	0.29	302.24	684.64	234.41	0.48	4178.33	314.87	2.41	19.08
22	Tambo, 35069	24.88°S; 124.57°E	395	0.24	307.08	704.12	244.29	0.48	4178.33	296.61	2.21	20.54
23	Tjukayirla, 13040	27.15°S; 124.57°E	450	0.25	313.24	664.66	242.10	0.51	4178.33	276.92	2.63	21.88

(continued on next page)

Table 1 (continued)

Site #	Name BOM ID	Geographic Inputs		Mean Annual Statistics MODIS Terra Satellite Input(s)				Target
		Location	Elevation (m)	NDVI	Land Surface Temperature (LST), (K)	Top Cloud Pressure (P_c , clouds) (Pa)	Top Cloud Temperature (T_c , clouds) (K)	
24	Townsville, 32040	19.25°S; 146.77°E	4	0.24	314.30	604.71	226.25	0.54
25	Wandana, 8294	28.21°S; 115.29°E	260	0.25	314.09	627.94	238.05	0.50
26	Youanmi, 12201	31.70°S; 119.24°E	420	0.19	311.08	596.44	226.56	0.56
15 Testing Sites								
ST1	Alice Springs, 15590	23.80°S; 133.89°E	546	0.23	311.82	667.99	242.86	0.51
ST2	Emerald Radar, 35146	23.55°S; 148.24°E	188	0.36	307.74	690.88	243.70	0.52
ST3	St George, 043053	28.05°S; 148.56°E	200	0.23	313.99	629.08	237.39	0.56
ST4	Marion Downs,	23.36°S; 139.66°E	124	0.63	290.06	661.75	226.31	0.69
ST5	Strathburn, 027032	14.48°S; 142.33°E	77	0.36	306.74	668.22	236.05	0.50
ST6	White Cliffs, 046093	30.67°S; 142.54°E	177	0.22	312.03	582.88	235.22	0.56
ST7	Nimmitabel, 070067	36.51°S; 149.28°E	1075	0.34	304.34	642.26	237.99	0.59
ST8	Belton, 19004	32.24°S; 138.71°E	485	0.36	314.88	704.44	249.65	0.51
ST9	Yalata Comm. 18161	31.48°S; 131.84°E	70	0.23	311.82	667.99	242.86	0.51
ST10	Lake Erella, 14515	12.50°S; 135.80°E	70	0.16	307.31	692.98	237.61	0.52
ST11	Tanami Downs, 15555	20.57°S; 129.73°E	305	0.37	312.62	621.64	226.25	0.50
ST12	Munarra, 7101	26.28°S; 118.69°E	500	0.32	303.92	665.41	232.29	0.56
ST13	Udalla, 3024	17.95°S; 123.4°E	20	0.36	303.04	687.90	235.22	0.63
ST14	Noondoonia, 11010	32.31°S; 123.73°E	170	0.56	306.46	610.86	222.73	0.48
ST15	Giles Met, 130170	25.03°S; 128.30°E	598	0.52	305.76	636.56	227.54	0.58

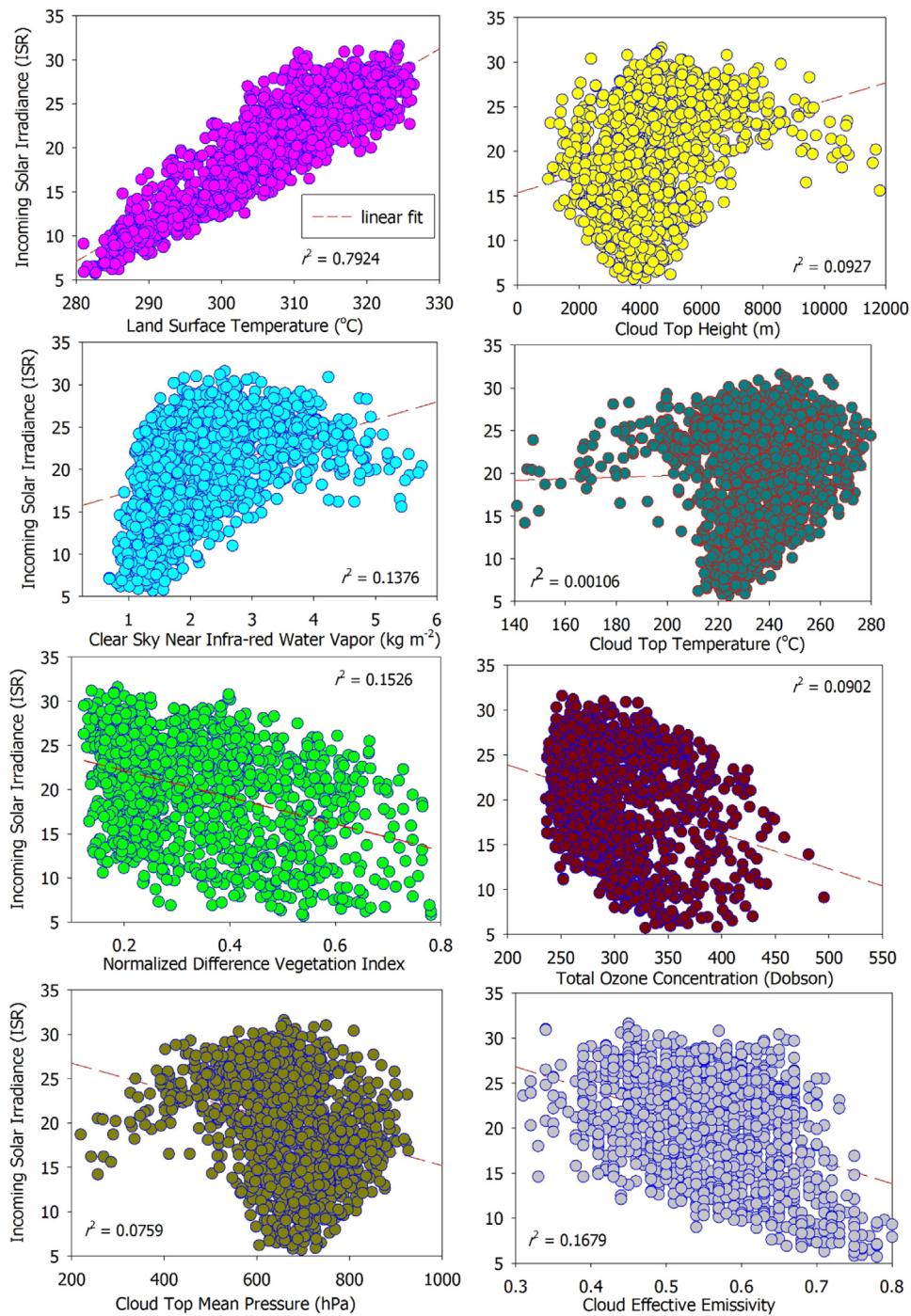


Fig. 3. Comparison of monthly *ISR* (i.e., the objective variable) with respect to satellite-derived predictor variables plotted in the training phase. A least square regression line with the coefficient of determination (r^2) is included for each sub-panel.

output layer contains the target *ISR*. In order to facilitate best feature extraction for a universal model, several neuronal activation functions (e.g., sine, hard limit, radial basis, triangular basis, logarithmic sigmoid, and tangent sigmoid; Eqs (11.1)–(11.6)) are trialled on the ELM architecture following earlier approaches [25,103]. The number of hidden neurons is varied from 1 to $n + 1$ (in increments of 1), where n is the number of training data points in each predictor-target matrix.

Based on the objective criterion, the mean square error for the trained model is monitored in each trial and the respective neuronal architecture is evaluated. All four years of data (2012–2015) are incorporated, however, the universal ELM model is designed with predictive features from 26 training sites across Australia (Fig. 2 and

Table 1). The pooled data from 26 sites enables a representative, regionally-adaptive, predictive model that is finally evaluated at the remaining 15 test sites. Since the ELM model requires random initialization of hidden layers, it is executed 1000 times to attain a suitable neuronal arrangement. The optimal neuronal architecture that generates the smallest *MSE* is selected, based on multi-model realisations of ELM (Table 2).

The RF model employs bootstrap aggregation ('bagging') to generate an ensemble of decision trees. Using the trees, RF is applied to regress the exploratory and response relationships between MODIS Terra satellite and geo-temporal inputs with respect to *ISR* data. A key parameter, the number of decision trees, is optimized by a trial of

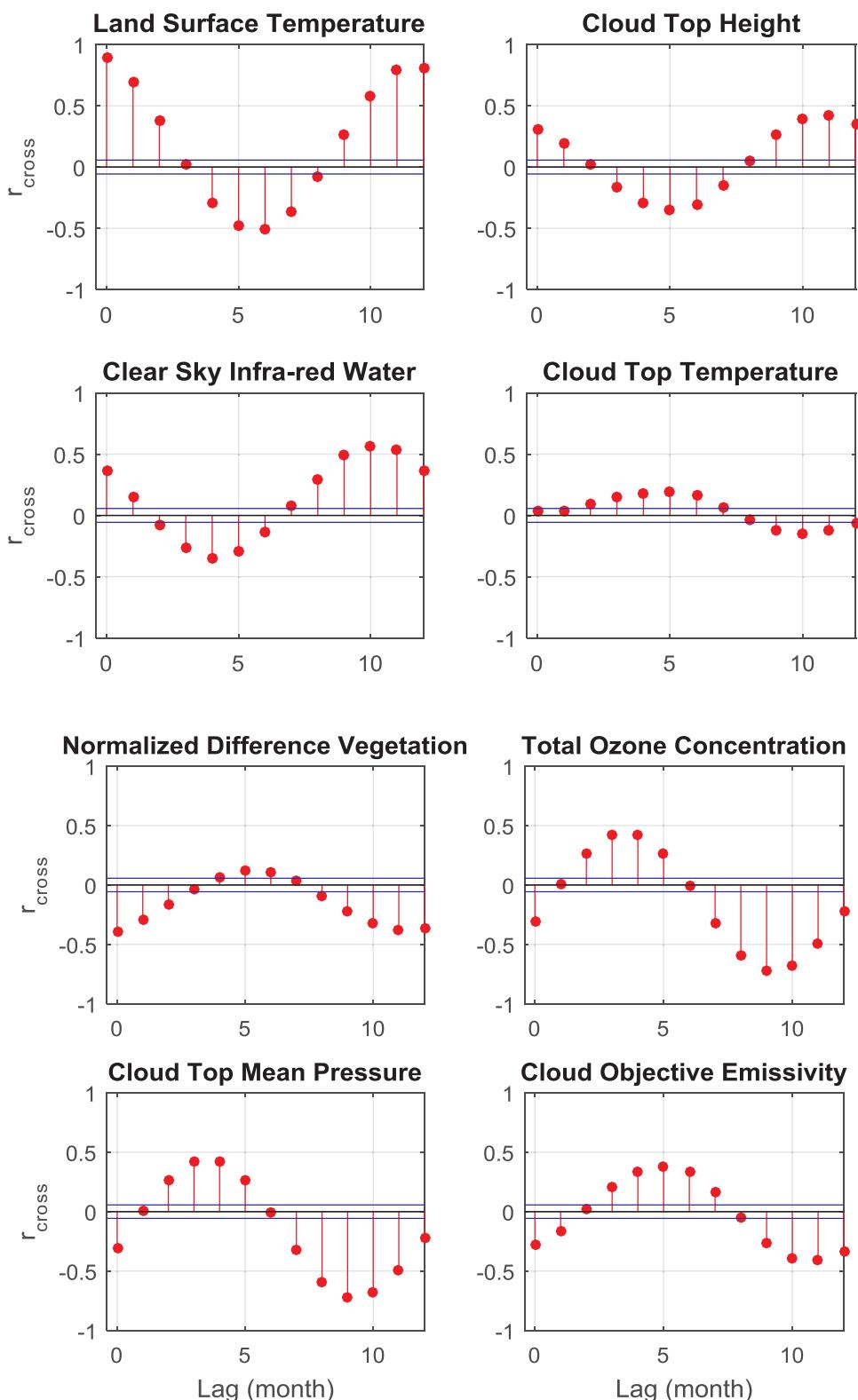


Fig. 4. Correlogram showing the covariance between the objective variable (IISR) and the predictor variable (satellite-variables) in terms of the cross-correlation coefficient (r_{cross}) in the model's training phase. Blue lines indicate the significance of r_{cross} at the 95% confidence interval for each panel.

several decision tree numbers (50–1600) in a two-fold increment. The optimal RF model is attained with 800 trees, a ‘leaf’ set to 5 and a ‘foot’ set to 1 (Table 3). In RF models the surrogate option is ‘on’ to allow predictive measures of variable association to fill each decision tree averaged over the entire split. Using bootstrap replicas of the satellite-based predictors for the 26 sites, every tree is grown independently and

then drawn to accord with the boot-strapping principle of resampled data with replacement. Observations not included in this replica are termed as “out of bag, OOB”. The error of the bagged ensembles is estimated by comparing each tree’s predictions based on its OOB averaged over the ensemble and the predicted OOB response with the true values. The split criterion and number of decision splits for each

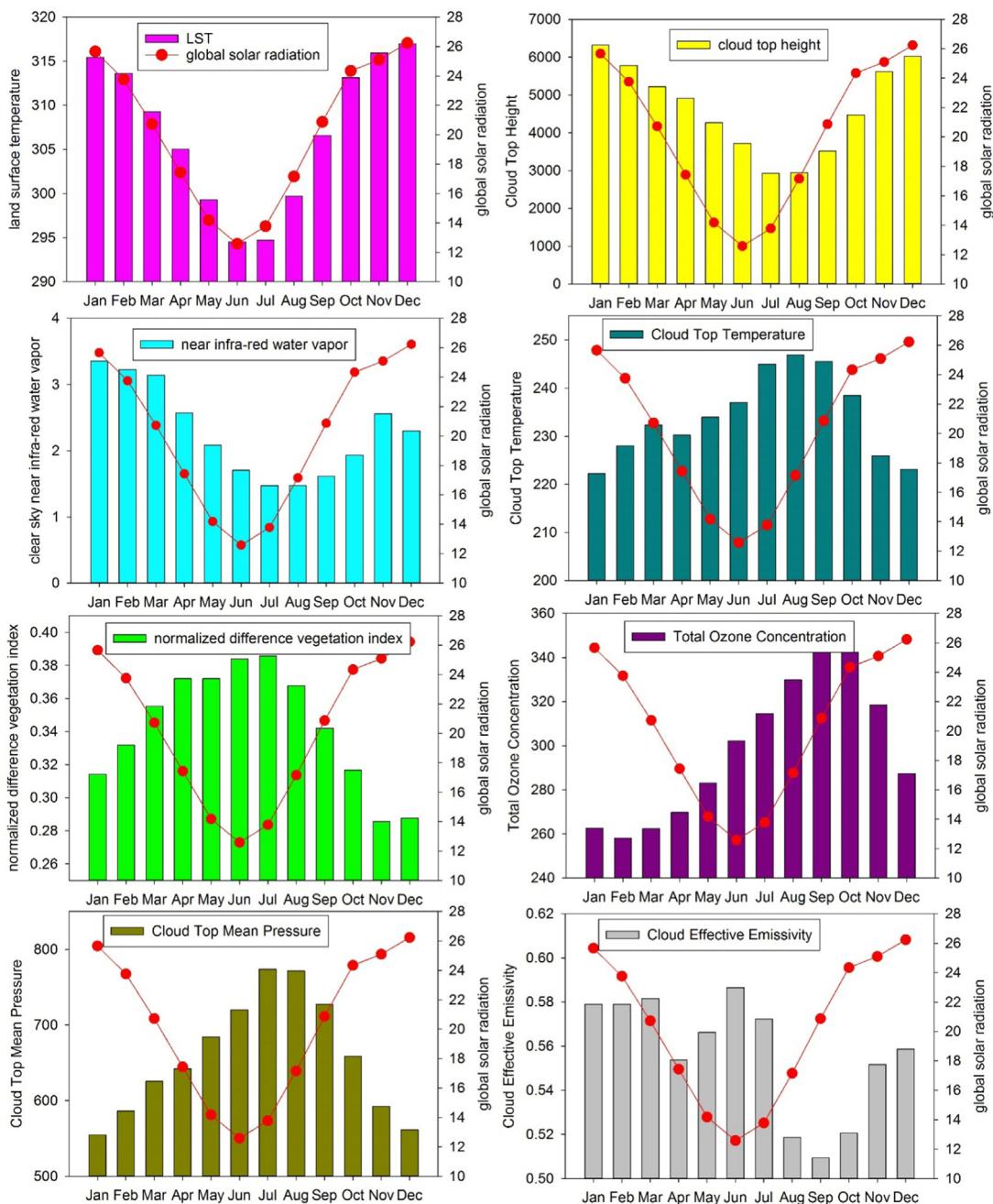


Fig. 5. Comparison of the monthly climatological cycle of satellite-derived predictor data and *ISR* averaged for the entire period 2012–2015.

predictor variable is found to be unique. Table 2 shows the optimal RF parameters, training error, and the correlation between observed and forecasted *ISR*.

The M5 Tree model is developed in consultation with the M5PrimeLab manual, M5' regression tree, and a model tree toolbox for Matlab/Octave (ver. 1.0.1) following Jekabsons [104]. To create an optimal model several training parameters for model initialization are specified where a tree split of 2.0, a smoothing level of 15, and a split threshold of 0.05 is applied based on earlier studies [102,103,105]. The option to prune the M5 Tree model tree is applied and a ‘divide-and-conquer rule’ is implemented [75,77]. The optimum decision-tree (or the ‘rules’ for the M5 Tree) is determined by the smallest RMSE and largest r^2 , concurring with the standard deviation of class values and the highest reduction in model training error. The M5 Tree model is fine-tuned to yield the most suitable set of rules. Table 2 shows the parameters of the trained model, including the r and RMSE.

A MARS model is constructed according to the ARESLab toolbox (version 1.13.0) [90], using the cubic function due to its superior smoothness over linear functions [106], and a better ability to model non-linearities between MODIS Terra satellite-based predictors and *ISR*. Given a better capacity to handle multiple satellite-based predictors, the generalized recursive partitioning regression (GRR) algorithm is employed as an adaptive tool for optimum function approximation [107].

Through a two-phase optimization (*i.e.*, forward and backward deletion process), the MARS algorithm initially ran with a ‘naïve’ model consisting of the intercept term, and thereafter, it iteratively added the reflected pair(s) of basis functions. The two-phase process yields the greatest decrease in training error (RMSE) where forward phase continues until one of these conditions is met:

- a. The maximum basis functions reached $\min[200, \max(20, 2d) + 1]$,

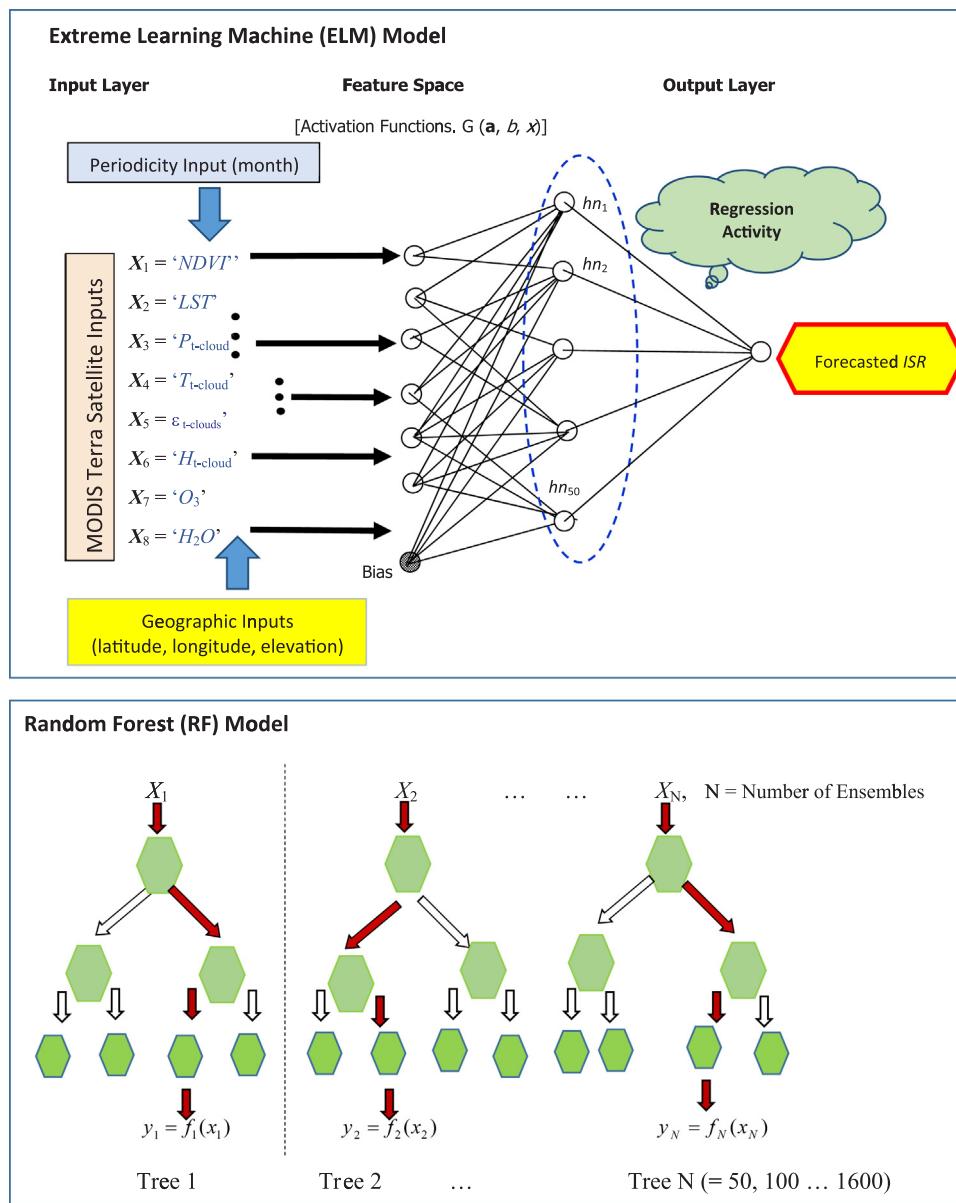


Fig. 6. Schematic diagram of ELM, RF, M5, and MARS models.

- where d = the number of inputs [88], or
- Adding a new basis function led to a change in the r^2 value by less than a threshold of 1×10^{-4} ,
 - The r^2 value reached approximately 1.0
 - The basis functions and the intercept term mimicked the data observations points
 - The number of effective parameters reached the total number of training data points.

An elaborately large MARS model generated at the end of the forward phase is likely to over-fit the training data as a number of splines are used to construct the predictive equation. Therefore, the net MARS model is pruned with a backward deletion phase, deleting redundant basis functions to reduce training error until the model includes only the intercept term. The MARS model is selected according to the lowest Generalized Cross-Validation (GCV) (Eq. (12)).

Table 2 shows model parameters and the equation (in training period) with r and RMSE.

3.4. Performance evaluation criteria

3.4.1. Mathematical formulations

To coincide with the formulations of the American Society for Civil Engineers [108], the model's performance evaluations are based on visual and descriptive statistics, where forecasted ISR is used to check its deviation from observed ISR for testing study sites. The mathematical equations of the performance metrics are as follows [109–111].

I. Correlation coefficient (r):

$$r = \left(\frac{\sum_{i=1}^N (ISR_i^{for} - \langle ISR_i^{obs} \rangle)(ISR_i^{for} - \langle ISR_i^{for} \rangle)}{\sqrt{\sum_{i=1}^N (ISR_i^{obs} - \langle ISR_i^{obs} \rangle)^2} \sqrt{\sum_{i=1}^N (ISR_i^{for} - \langle ISR_i^{for} \rangle)^2}} \right) \quad (21)$$

II. Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (ISR_i^{for} - ISR_i^{obs})^2} \quad (22)$$

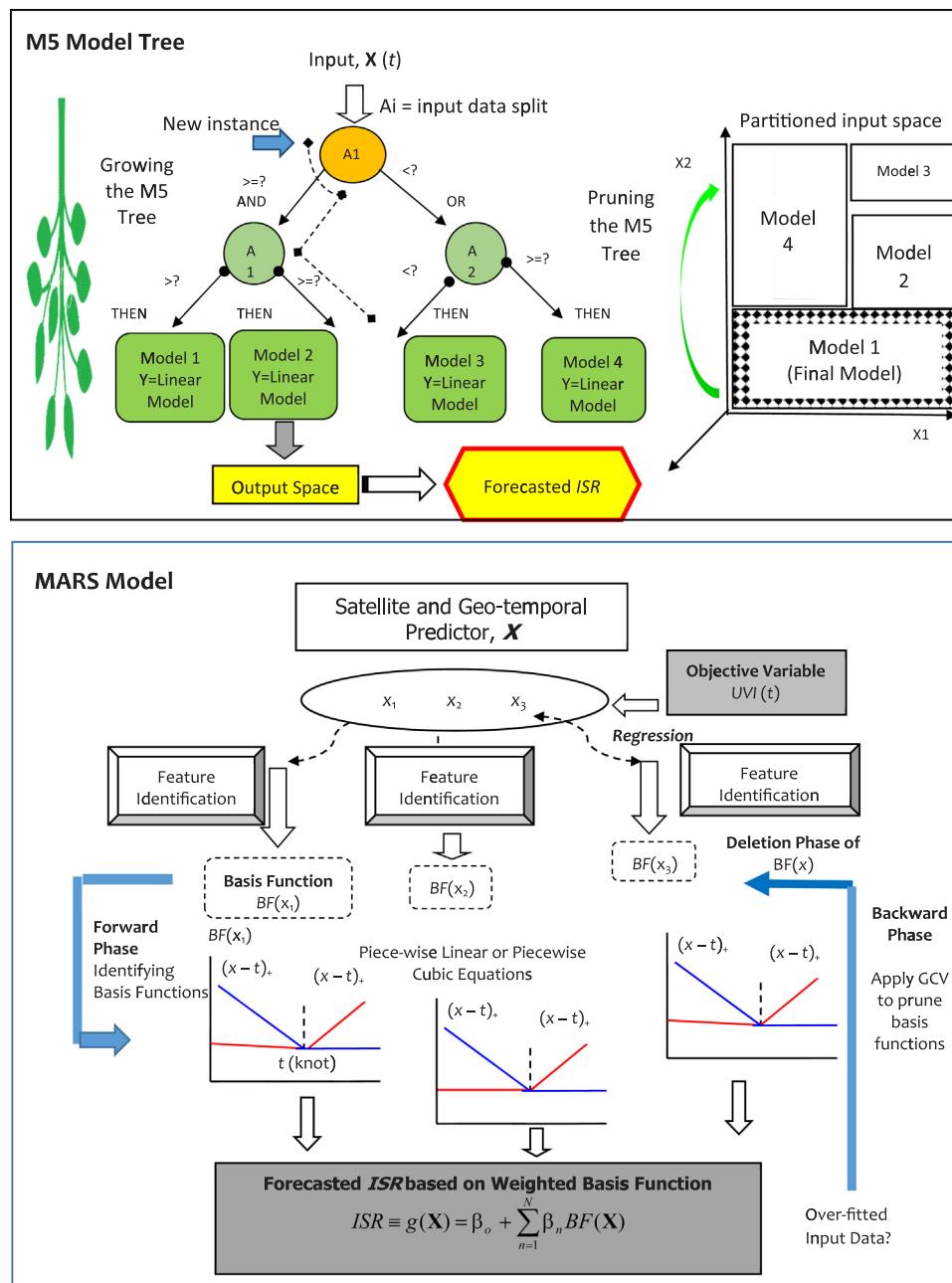


Fig. 6. (continued)

III. Mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |(ISR_i^{for} - ISR_i^{obs})| \quad (23)$$

IV. Willmott's Index (WI):

$$WI = 1 - \left[\frac{\sum_{i=1}^N (ISR_i^{for} - ISR_i^{obs})^2}{\sum_{i=1}^N (|ISR_i^{for} - \langle ISR_i^{obs} \rangle| + |ISR_i^{obs} - \langle ISR_i^{obs} \rangle|)^2} \right], \quad 0 \leq WI \leq 1 \quad (24)$$

V. Nash-Sutcliffe coefficient (E_{NS}) [109]:

$$NS = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \right], \quad -\infty \leq NS \leq 1 \quad (25)$$

VI. Legates and McCabe Index (E_{LM}) [112]:

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^N ISR_i^{obs} - ISR_i^{for}}{ISR_i^{obs} - \bar{ISR}_i^{obs}} \right], \quad 0 \leq E_{LM} \leq 1 \quad (26)$$

VII. Relative root mean square error (RRMSE %):

$$RRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (ISR_i^{for} - ISR_i^{obs})^2}}{\frac{1}{N} \sum_{i=1}^N (ISR_i^{obs})} \times 100 \quad (27)$$

VIII. Mean absolute percentage error (MAPE; %):

Table 2

ELM and comparative model's (*i.e.*, RF, M5 Model Tree and MARS) training performances. r = correlation coefficient and RMSE = root mean square error in training period computed between the observed and forecasted ISR data.

Model	Design Parameters	Training Performance	
		r	RMSE
ELM	Number of Layers	3	0.9790
	Input Neurons	12	
	Inputs	[geographic, periodicity & satellite-based data]	
	Hidden Neurons	10, 20, ..., 150	
	Output Neurons	1 (global solar radiation, ISR)	
	Activation Functions	<i>Sigmoid, sine, hard limit, triangular basis, radial basis, tangent sigmoid, logarithmic sigmoid.</i>	
RF	Optimal Architecture	12 – 110 – 1 (input-hidden-output) determined iteratively by trial and error	1.5469
	Number of Trees	50, 100, 200, 400, 800, 1600 (optimal value 800)	
	Leaf	5	
	Foot	1	
M5 Tree	Surrogate	'on'	1.3528
	Number of Rules	147	
	Minimum Cases	5	
MARS	Smoothing	15	1.2949
	Split Threshold	0.05	
	Number of Basis Functions	18	
MARS	Spline Type	'cubic'	0.9756
	Generalized Cross Validation	1.7751	
	Model Equation	$y = -1.01e + 03 + 0.0988*BF1 - 0.36*BF2 + 0.653*BF3 + 2.83$ $* BF4 - 0.009*BF5 + 0.552*BF6 - 0.0768*BF7 - 8.47$ $* BF8 - 1.64*BF9 + 1.25*BF10 + 2.32*BF11 + 0.000749$ $* BF12 + 0.0102*BF13 - 0.13*BF14 - 946*BF15 + 943$ $* BF16 + 946*BF17'$	

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{(ISR_i^{for} - ISR_i^{obs})}{ISR_i^{obs}} \right| \times 100 \quad (28)$$

where ISR^{obs} and ISR^{for} are the observed and forecasted i th value of ISR, $\langle ISR \rangle^{obs}$ and $\langle ISR \rangle^{for}$ are the observed and forecasted mean G in the cross-validation (testing) phase and N is the number of data points.

The study has also used $RMSE_{ss}$ as an additional skill score metric that is benchmarked with a persistence type model ($RMSE_p$) in which the immediate antecedent (past) value is used to estimate current value, with respect to measured global incident solar radiation as a reference.

IX. Skill Score Metric ($RMSE_{ss}$) [27]

$$RMSE_{ss} = \frac{RMSE^f}{RMSE^p} \quad (29)$$

3.4.2. Physical interpretation

The physical interpretation of performance metrics in the context of the problem is as follows. The evaluation of ELM (*vs.* RF, M5 Tree, and MARS models) utilizes Pearson's product-moment correlation coefficient bounded by $[0, 1]$, where 0 corresponds to a relatively poor model and 1.0 corresponds to a perfect model. The value of r , or its equivalent R^2 (*i.e.*, the coefficient of determination), is a first-order metric describing the quantity of variance (or alternatively the degree of collinearity) in observed ISR that can be explained by the respective solar energy simulation model [111,113]. Considering several criticisms (*e.g.*, [114,115]), the r metric may not fundamentally indicate cases where a strong agreement exists between model and observations, since it is based on linear agreement. While correlation statistics are thus acceptable ways to assess a model, performance metrics in the units of the respective variable (*e.g.*, $MJ\ m^{-2}\ day^{-1}$) are required for many practical applications and physical interpretations. This can be particularly useful for study sites with similar solar radiation potential. The solar energy assessment model used may generate differing accuracy levels that then need to be quantified in their absolute terms. Moreover, the standardizations inherent in correlation-based measures (*e.g.*, r or r^2) can be considered as limiting metrics in the sense that a high

correlation can be achieved by a mediocre or even a poor solar radiation predictive model [112].

The absolute errors (*i.e.*, RMSE and MAE) are able to provide better information about a model's predictive skill. RMSE quantifies the goodness-of-fit relevant to high ISR values (particularly when error distribution is Gaussian) whereas MAE is not weighted towards higher magnitude or lower magnitude but instead evaluates all deviations from observed data in an equal manner regardless of the sign (*i.e.*, when error distribution is normal) [113]. Due to a slight bias (*i.e.*, inflated values) in RMSE when large outliers are present, the MAE is slightly preferred. However RMSE is important to quantify differences and to indicate the extent to which outliers (or variances between forecasted and observed ISR) exist (when error distribution is Gaussian) [113]. RMSE can be biased in favour of the peaks and higher magnitudes that will in most cases exhibit the greatest error, and are insensitive to the lower magnitude recorded data [111]. Although RMSE satisfies the triangle inequality for a distance function metric [113], and therefore must not be totally discounted, it is also likely to be more sensitive than other performance metrics (including MAE) due to occasionally elevated model errors as the squaring of the terms may yield disproportionate weight to the larger errors [112]. Indeed, Willmott *et al.* [116] concluded that the sums-of-squares based error statistics (*e.g.*, RMSE) may have inherent ambiguities over alternates such as MAE. Consequently, both RMSE and MAE must be used for model evaluation to obtain accuracy in absolute units.

While correlation and absolute measures (*e.g.*, MAE and RMSE) are relevant metrics, at the crux of model evaluation in geographically diverse sites (such as the spatial-scale model evaluation undertaken in this study), is the requirement for relative percentage errors comparable from one site to another. To provide a universally quantifiable metric the Willmott's Index (WI) and Nash-Sutcliffe's efficiency (E_{NS}) [109,115,117] are equally important. WI considers the ratio of the mean square error instead of the square of the error differences [117] and can provide an advantage over r , RMSE, and MAE. In the physical sense, E_{NS} represents the mean square error in the model corresponding to the variance in the observed incident solar radiation, subtracted from a reference for a perfect model (*i.e.*, unity). If the squared difference between forecasted and observed ISR is relatively large to concur with the variance in actual ISR values, E_{NS} will be zero (depicting a generally

Table 3

Evaluation of ELM vs. the RF, M5 Tree and MARS models at 15 test sites assessed with the correlation coefficient (r), root mean square error ($RMSE$; MJ m $^{-2}$), mean absolute error (MAE ; MJ m $^{-2}$) and the standardized performance metrics (Willmott's Index, WI & Nash-Sutcliffe coefficient, E_{NS}) between the forecasted and observed ISR data in the testing phase. The best performance site is marked **blue** (White Cliffs), the worst site is marked **red** (Lake Evella), and a skill score metric based on the forecasted $RMSE$ relative to a persistence model error, $RMSE_{ss}$ is included.

Site	ELM						RF					
	r	$RMSE$	MAE	$RMSE_{ss}$	WI	E_{NS}	r	$RMSE$	MAE	$RMSE_{ss}$	WI	E_{NS}
ST1	0.953	1.383	1.038	0.554	0.955	0.896	0.943	1.569	1.296	0.494	0.949	0.866
ST2	0.984	0.780	0.596	0.709	0.983	0.966	0.975	1.009	0.809	0.624	0.970	0.942
ST3	0.980	1.054	0.794	0.657	0.980	0.959	0.968	1.297	1.019	0.578	0.969	0.937
ST4	0.982	0.866	0.667	0.715	0.983	0.963	0.974	1.290	1.080	0.575	0.953	0.919
ST5	0.888	1.384	1.031	0.575	0.867	0.788	0.780	1.908	1.518	0.414	0.729	0.596
ST6 (most accurate)	0.993	0.752	0.576	0.787	0.993	0.985	0.988	0.993	0.783	0.716	0.987	0.974
ST7	0.991	0.875	0.665	0.773	0.990	0.981	0.968	2.845	2.192	0.261	0.911	0.801
ST8	0.994	0.792	0.617	0.785	0.994	0.986	0.990	1.013	0.729	0.728	0.989	0.977
ST9	0.988	1.033	0.843	0.702	0.987	0.972	0.982	1.381	1.118	0.602	0.973	0.951
ST10 (least accurate)	0.912	1.498	1.140	0.495	0.799	0.761	0.720	2.246	1.509	0.242	0.467	0.462
ST11	0.944	1.131	0.872	0.572	0.949	0.890	0.922	1.346	1.069	0.491	0.931	0.844
ST12	0.990	0.828	0.597	0.745	0.988	0.976	0.980	1.141	0.954	0.649	0.979	0.954
ST13	0.946	0.982	0.780	0.653	0.949	0.892	0.918	1.320	1.098	0.533	0.875	0.805
ST14	0.987	1.157	0.821	0.691	0.983	0.965	0.983	1.355	1.126	0.638	0.973	0.952
ST15	0.972	1.078	0.857	0.677	0.976	0.944	0.959	1.296	1.060	0.612	0.965	0.920
All Site-Average	0.967	1.040	0.793	0.673	0.958	0.928	0.937	1.467	1.157	0.544	0.908	0.860
M5 Tree												
Site	r	$RMSE$	MAE	$RMSE_{ss}$	WI	E_{NS}	r	$RMSE$	MAE	$RMSE_{ss}$	WI	E_{NS}
ST1	0.917	1.910	1.557	0.384	0.926	0.801	0.945	1.726	1.309	0.443	0.940	0.837
ST2	0.949	1.387	1.110	0.482	0.938	0.891	0.976	1.052	0.787	0.607	0.971	0.937
ST3	0.941	1.764	1.343	0.426	0.945	0.884	0.979	1.080	0.811	0.649	0.979	0.956
ST4	0.932	2.074	1.545	0.317	0.844	0.790	0.969	1.144	0.868	0.623	0.971	0.936
ST5	0.710	2.294	1.751	0.296	0.685	0.417	0.787	2.750	2.085	0.156	0.508	0.162
ST6	0.973	1.439	1.044	0.589	0.974	0.946	0.990	0.929	0.727	0.734	0.990	0.978
ST7	0.939	4.027	3.409	-0.046	0.864	0.600	0.980	1.747	1.528	0.546	0.967	0.925
ST8	0.986	1.155	0.902	0.689	0.987	0.970	0.992	0.992	0.773	0.733	0.990	0.978
ST9	0.969	1.865	1.385	0.462	0.952	0.910	0.987	1.096	0.878	0.684	0.984	0.969
ST10	0.697	2.326	1.452	0.215	0.453	0.422	0.661	3.821	2.853	-0.289	0.228	-0.558
ST11	0.880	1.657	1.299	0.373	0.897	0.764	0.195	10.106	2.580	-2.824	0.218	-7.773
ST12	0.971	1.329	1.085	0.591	0.972	0.938	0.984	1.098	0.871	0.662	0.981	0.958
ST13	0.739	2.189	1.708	0.226	0.564	0.463	0.842	1.717	1.271	0.393	0.857	0.670
ST14	0.971	1.651	1.352	0.559	0.960	0.929	0.991	1.022	0.822	0.727	0.985	0.973
ST15	0.952	1.467	1.189	0.560	0.959	0.897	0.967	1.249	0.949	0.626	0.970	0.925
All Site-Average	0.902	1.902	1.475	0.408	0.861	0.775	0.883	2.102	1.274	0.298	0.806	0.192

poor model). If this metric is negative, the results indicate that the observed mean is a better predictor than the model (depicting a very poor model) [118]. With a value of 1.0, E_{NS} is likely to indicate a perfect model [112]. In fact, some studies (e.g., [119]) report that even for poorly-fitted models, the E_{NS} can be relatively high (> 0.8), while the best models may not generate values that are acceptably high. The insensitivity of E_{NS} is overcome, to some degree, by WI [117], where the ratio of mean square to potential error, multiplied by the number of test points, is subtracted from unity. However, a certain degree of insufficiency still exists with WI , due to the squaring of the denominator terms.

To overcome this insufficiency, Legate & McCabe's Index (E_{LM}) is used as a "baseline adjusted" metric where errors and differences are given their appropriate weighting and are not inflated by squared values. This index also varies from 0 to 1, with higher values indicating a better fit of the model. Since geographic differences between study sites (Table 1; Fig. 2) can lead to differences in solar radiation distribution, the relative root mean square error ($RRMSE$) is also computed [23,120] to compare the model over geographically diverse sites. According to [121], a solar radiation model's precision is excellent if the $RRMSE < 10\%$, good if $10\% < RRMSE < 20\%$, fair if $20\% < RRMSE < 30\%$

and poor if the $RRMSE > 30\%$. The ratio of the $RMSE$ to a persistence model error is particularly important to provide a baseline reference comparison. However, the interpretation of a universal model must be made in conjunction with the WI , E_{NS} , and E_{LM} including absolute measures ($MAE/RMSE$) and visual measures (e.g., Taylor diagrams, statistical moments of errors, such as mean, variance, skewness, and flatness, etc.) to yield a complete assessment of the model's error variation [113,122].

4. Results and discussion

The spatially-universal ELM-based predictive model developed with MODIS Terra satellite predictors used training data from 26 study sites and was evaluated with an independent set of 15 test study sites over the period 2012–2015. Fig. 7 is a scatterplot of forecasted vs. observed ISR at all 15 cross-validation sites. In each panel, the least square regression line and the coefficient of determination (r^2) with a linear fit equation $y = mx + C$ describes the exactness of the forecasted ISR compared to real ISR values.

The degree of overall scatter is much less for the ELM compared to the RF, MARS, and M5 Tree models. This is corroborated by the r^2 -value

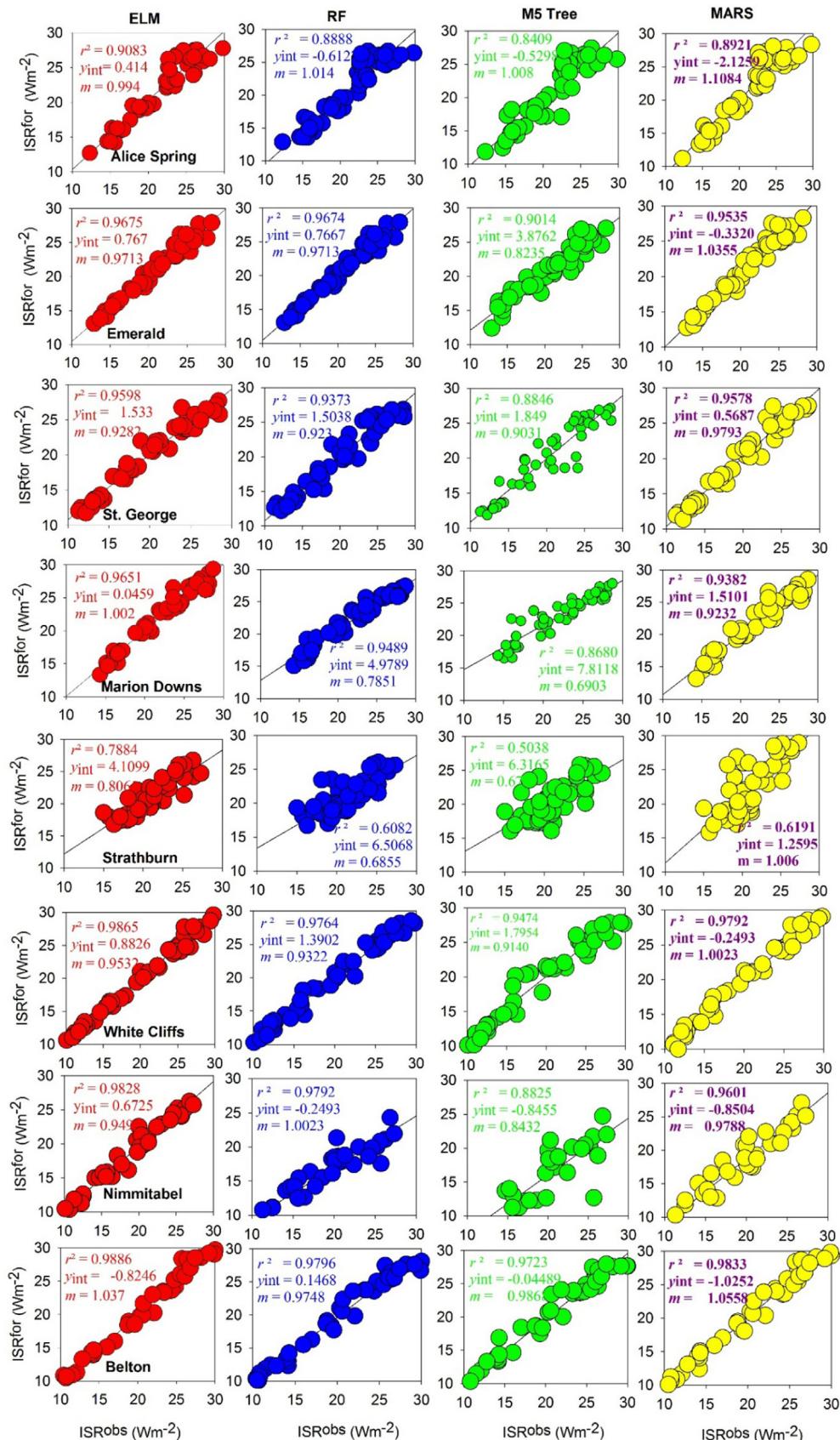


Fig. 7. Scatterplots of forecasted vs. observed ISR for the 15 tested sites across Australia. A least square regression line and coefficient of determination (r^2) with a linear fit equation $y = mx + C$ is shown in each sub-panel.

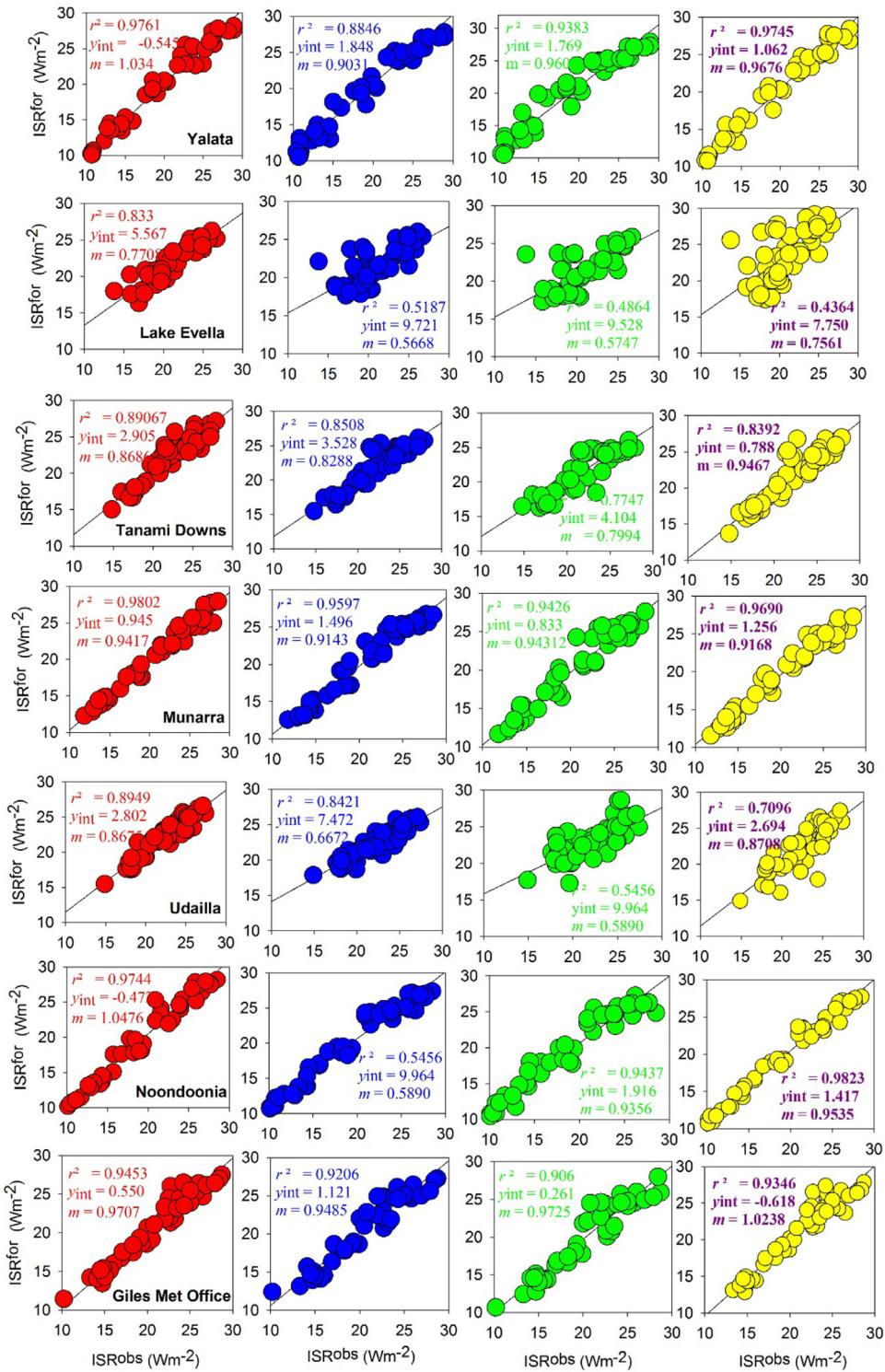


Fig. 7. (continued)

which is relatively high for the ELM model (greater than 0.90 for 11 of 15 cross-validation sites). The highest r^2 value for the ELM model is found for Belton ($r^2 \approx 0.9886$), followed by White Cliffs ($r^2 \approx 0.9865$), and Nimmitabel ($r^2 \approx 0.9828$). The RF model, however, has an $r^2 \approx 0.9796$, 0.9764, and 0.9792, the M5 Tree an $r^2 \approx 0.9723$, 0.8825, and 0.9474, and the MARS model an $r^2 \approx 0.9833$, 0.9792, and 0.9601 for the Belton, White Cliffs, and Nimmitabel sites, respectively. The high performance of the ELM model compared to the benchmark models is justified by the 1:1 correspondence between ISR^{for} and ISR^{obs} tested data. For example, the gradient of the linear fit is close to unity

with $m \approx 1.037$ compared to $m \approx 0.9748$ (RF), 0.9863 (M5 Tree), and 1.0558 (MARS) for one of the best performing sites (*i.e.*, ST8, Belton). Superior performance of the ELM model is also found for the poorest performing study site (*i.e.*, ST5, Strathburn) among 15 tested sites, where the r^2 (0.7874) attained by the ELM model is notably larger than r^2 values of 0.6082 (RF), 0.5038 (M5 Tree), and 0.6191 (MARS). When translated to a correlation coefficient, the results for Strathburn show that at least 88.73% of the inherent covariance in the forecasted ISR can be explained by observed ISR for the ELM model, whereas only 77.98% is explained by RF, 70.97% by M5 Tree, and 78.68% by the MARS

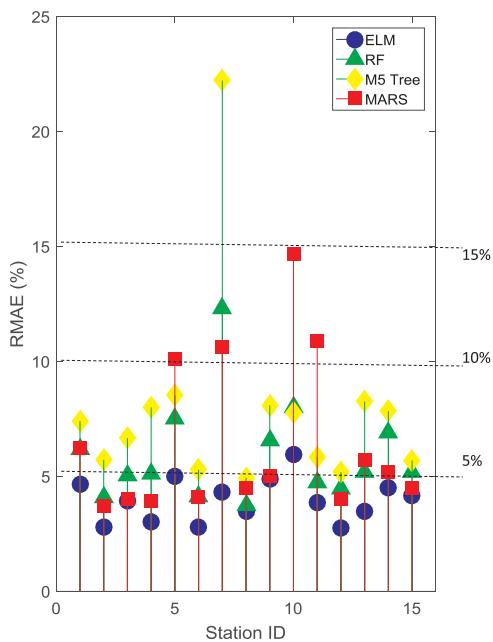


Fig. 8. Stair plot of the relative mean absolute error (RMSE, %) generated by the ELM vs. the RF, M5 Model Tree, and MARS models applied at the 15 tested sites. Station ID corresponds to the Site # in Table 1.

model. The results clearly indicate the superior quality of the universal ELM model to forecast global incident solar radiation.

A stair plot showing the *RMAE* generated by the ELM vs. the RF, M5 Model Tree, and MARS models at the 15 test study sites is shown in Fig. 8. Similar to correlation-based metrics, the percentage error values (in accordance with the mean absolute error) are significantly lower for the ELM model than for other data-driven models across all 15 study sites. Most importantly, the error for the ELM model is less than 5% for all but two sites, which is considered to be a very low forecasting error, and consequently indicates a high-performance model (e.g., [7,23,120,123]). Although the RF model also generates less than 5% error at a number of sites, error values are always higher than the ELM model. For M5 Tree and MARS models, the percentage errors are 5–10% higher than the ELM model. At stations 5, 7, 10, and 11, the errors exceed a value of 10% for *ISR* generated by the MARS model. The two study sites with the worst performance (ST5, Strathburn and ST10, Lake Evella), have an error generated by the MARS model of about 10.9% and 11.2%, respectively, whereas the error for the ELM model is only 6.6% and 7.2%. Similarly, poor performance of the RF and M5 Tree models relative to the ELM model is noted with respect to ELM for these worst performing study sites.

In Table 3, the study evaluates the ELM vs. RF, M5 Tree, and MARS models at the 15 test study sites using: r , $RMSE$ ($MJ\ m^{-2}$), MAE ($MJ\ m^{-2}$), WI , and E_{NS} . $RMSE_{SS}$ is used to compare the present models with a baseline persistence model [27]. The sites with optimal performance are marked in blue and the worst performing sites are marked in red.

Similar to Figs. 7 and 8, the performance metrics reveal the superior

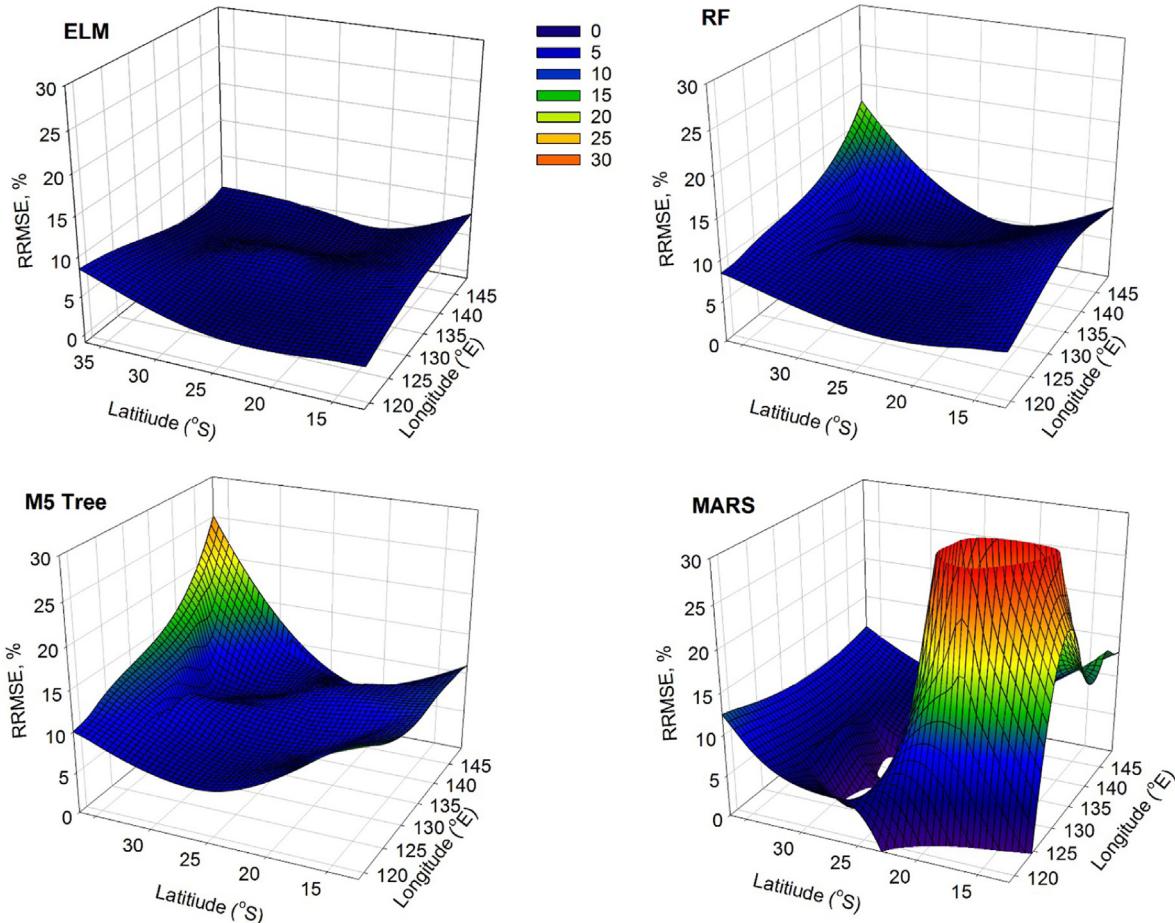


Fig. 9. Evaluation of ELM vs. the RF, M5 Tree, and MARS models in terms of their relative *RRMSE* error (%) presented as surface plots contoured over 15 test sites across Australia. Note: a lower *RRMSE* indicates a better forecasting accuracy.

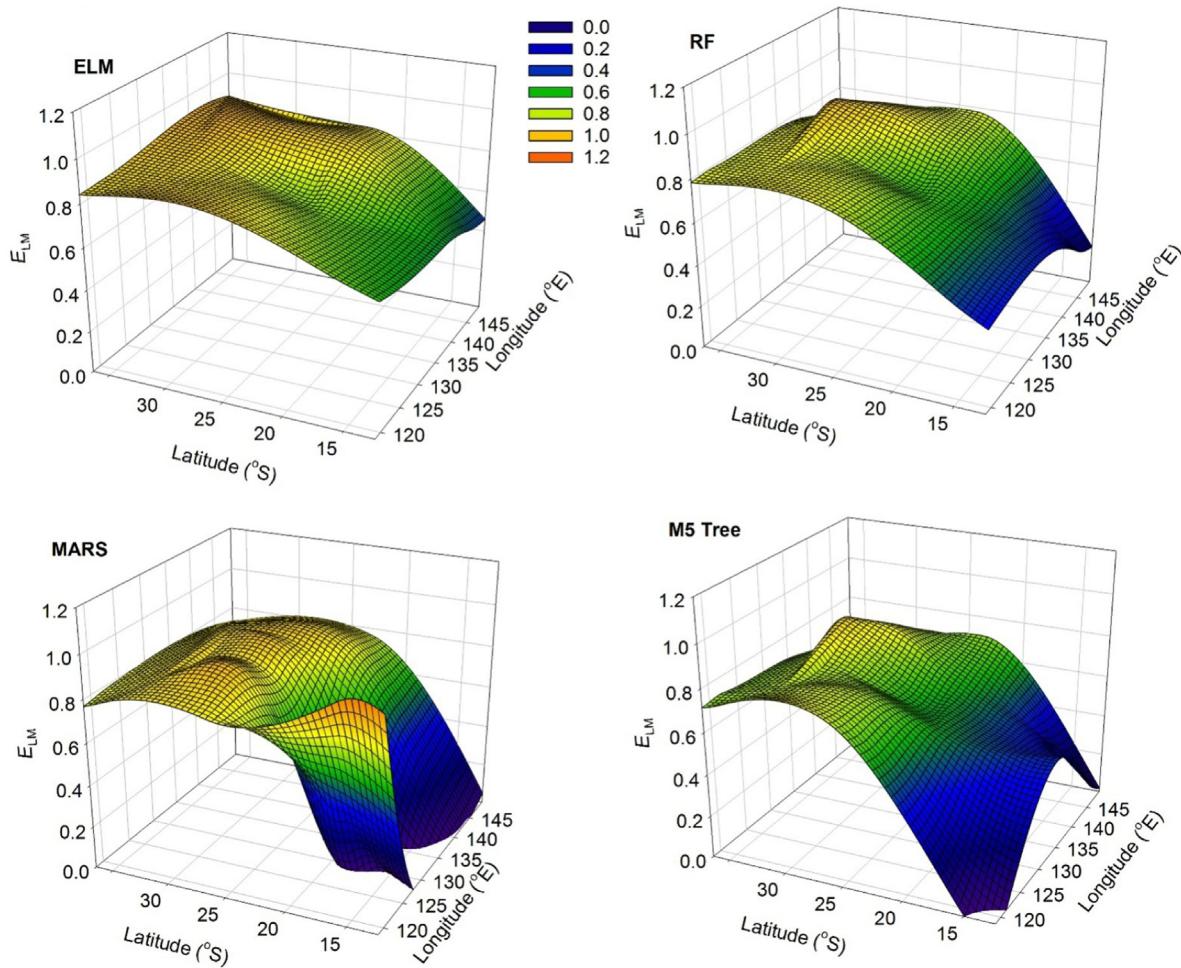


Fig. 10. Evaluation of ELM vs. RF, M5 Tree, and MARS models in terms of Legates & McCabe's Index (E_{LM}), presented as surface plots contoured over the 15 test sites across Australia. Note: a higher E_{LM} indicates a greater forecasting accuracy.

skill of ELM at all 15 test study sites, albeit, with a notable geographic signature (Table 3). $RMSE$ is significantly lower for ELM ($0.752 \leq RMSE \leq 1.492 \text{ MJ m}^{-2}$) compared to the RF model ($0.993 \leq RMSE \leq 2.246 \text{ MJ m}^{-2}$), the M5 Tree model ($1.155 \leq RMSE \leq 2.326 \text{ MJ m}^{-2}$), and the MARS model ($0.727 \leq RMSE \leq 2.853 \text{ MJ m}^{-2}$). High range errors are also recorded by all but the ELM model in terms of the MAE values for all 15 study sites. Of particular interest to the present error discussion is the skill score metric (i.e., denoting a comparison with a simple persistence-based model) registered by the ELM, which far exceeded the values obtained by the RF, M5 Tree, and MARS models.

At the most accurate study site (ST6, White Cliffs), the $RMSE_{SS}$ (≈ 0.787) for the ELM model is relatively high in contrast to $RMSE_{SS} \approx 0.783$, 0.589, and 0.784 obtained from the RF, M5 Tree, and MARS models, respectively. For the least accurate study site (ST10, Lake Evella) the ELM model has an $RMSE_{SS} \approx 0.495$, which is more than two-fold larger than the $RMSE_{SS}$ generated by RF, M5 Tree, and MARS (0.242, 0.215 and –0.289, respectively). For the Strathburn site (ST5), $RMSE_{SS}$ is relatively high, at about 0.575 for ELM compared to 0.414 (RF), 0.216 (M5 Tree) and 0.156 (MARS), which concurs with the correlation-based metrics (Fig. 7).

The skill score for ELM is more than 50% for 12 out of 15 tested study sites, whereas only 10 study sites for the RF and MARS model, and 6 study sites for the M5 Tree, have scores greater than 50%. Also, the values of WI and E_{NS} generated by the ELM model are closer to unity than those of RF, M5 Tree, and MARS, and the site-averaged statistical

metrics across Australia indicates that the ELM model has the best performance relative to its benchmark models.

In Figs. 9 and 10 the study evaluates ELM vs. RF, M5 Tree, and MARS models in terms of $RMSE$ and E_{LM} , presented as a set of 3-dimensional surface plots contoured over the 15 test study sites.

There is explicit evidence that the ELM model trained with MODIS Terra satellite data is able to generate very low relative percentage errors (i.e., $RRMSE < 5\%$) distributed spatially across Australia. For RF and M5 Tree models most of the tested study sites also reveal less than a 5% forecasted error, except in the southeast corner of Australia (i.e., ST 7, Nimmitabel) where $RRMSE$ reaches 15% and 20%, respectively. In spite of the good performance for these two models across Australia, there are subtle, yet notable differences (and generally larger errors) compared to the ELM model. Importantly, when the results for the MARS model are evaluated across Australia, significantly large forecasted errors are evident in the far eastern and southern parts of the Australia. These errors are as high as 30% and show generally poor performance of the MARS model when used for ISR forecasting with a limited set of MODIS Terra satellite-based predictor variables.

In Fig. 10, the performance metric (E_{LM}), contoured over Australia, is high and close to unity for the ELM in the majority of tested study sites. The RF model also performed generally well (i.e., $E_{LM} > 0.60$) except at the tested study sites located below 20° latitude where the E_{LM} metric is lower than 0.40. The performances of M5 Tree and MARS are worse than RF. Taken together with the results of Fig. 9, the very poor performance based on the E_{LM} metric indicates that the M5 Tree and

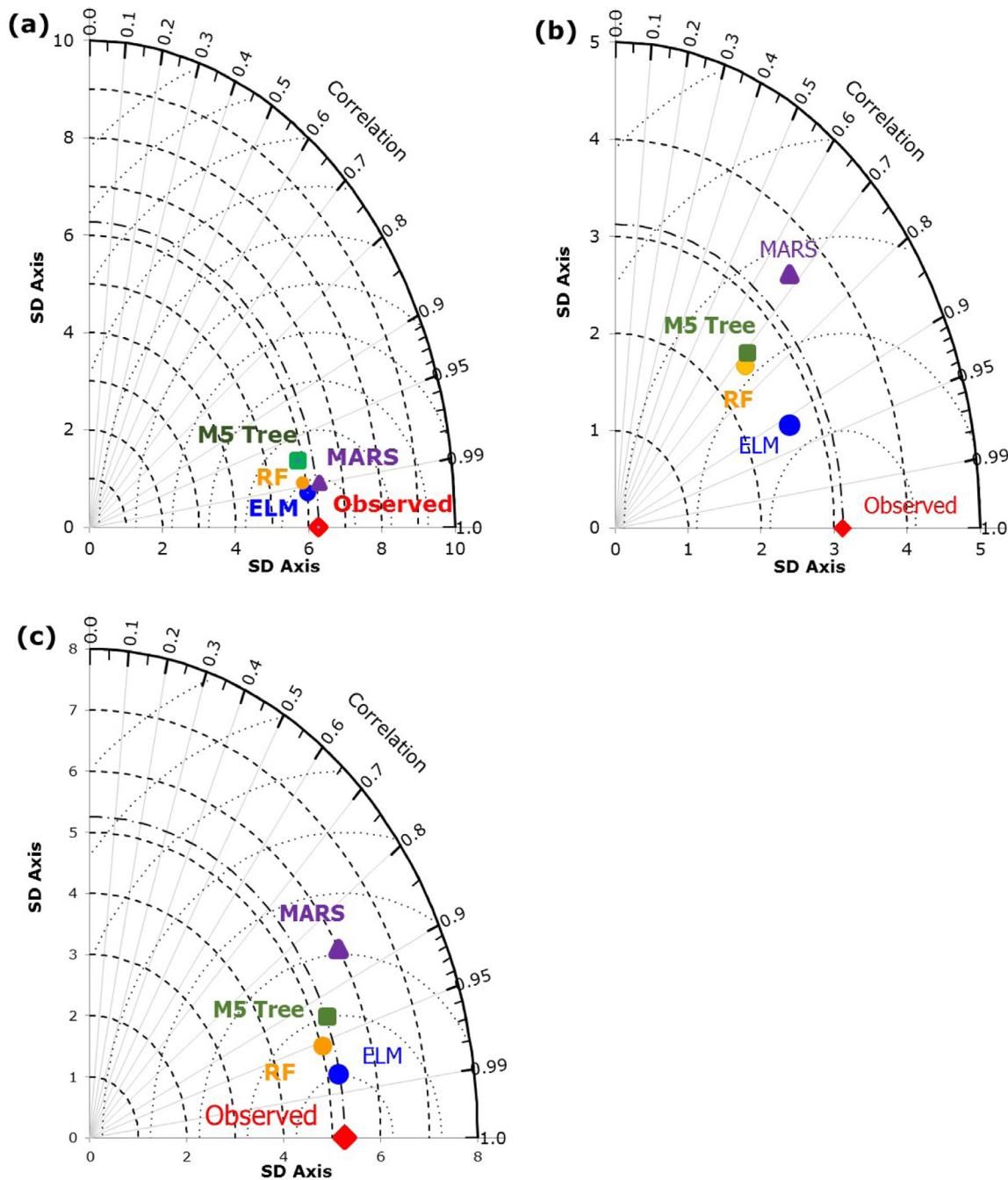


Fig. 11. Taylor diagram depicting the correlation coefficient and root-mean square centred difference when applying the ELM vs. the RF, M5 Tree, and MARS models for: (a) most accurately forecasted station (*i.e.*, ST 6, White Cliff), (b) least accurately forecasted station (*i.e.*, ST 10, Lake Evella), (c) all 15 test stations averaged over Australia.

MARS models can be discounted for large-scale spatial prediction of *ISR* using satellite predictors, and the objective ELM model is preferred over its three benchmark models.

Fig. 11(a)-(c) gives an overview of the correlation coefficients of forecasted and observed data including standard deviations and root-mean-square centred differences generated by ELM vs. the MARS, RF, and M5 Tree models. The figure shows the most accurate study site (ST6, White Cliffs), the least accurate study site (ST10, Lake Evella), and an average of all 15 tested study sites. This comparison corresponds with the previously indicated superior performance of the ELM over alternative predictive models, and the difference in performance level is most evident with the ‘all-station’ average and the least accurately

forecasted study site (Fig. 11(b)-(c)). For White Cliffs, the only non-acceptable performance is the M5 Tree (Fig. 11(a)). For the least accurate site and all-site averaged scenarios, the MARS model leads to a significantly inferior performance (Fig. 11(b)-(c)). The performances of the RF and M5 Tree are better than MARS, although the Taylor plot reveals that the RF model exceeds the other comparative models by a notable margin. Integrating several decision trees can help overcome the problem of overfitting, exhibit less variance than a single tree (*e.g.*, M5 Tree), and maintain accuracy even with missing data [69,124]. This is possibly the reason why RF outperforms the M5 Tree and MARS models, although ELM remains superior to all alternative models for *ISR* prediction.

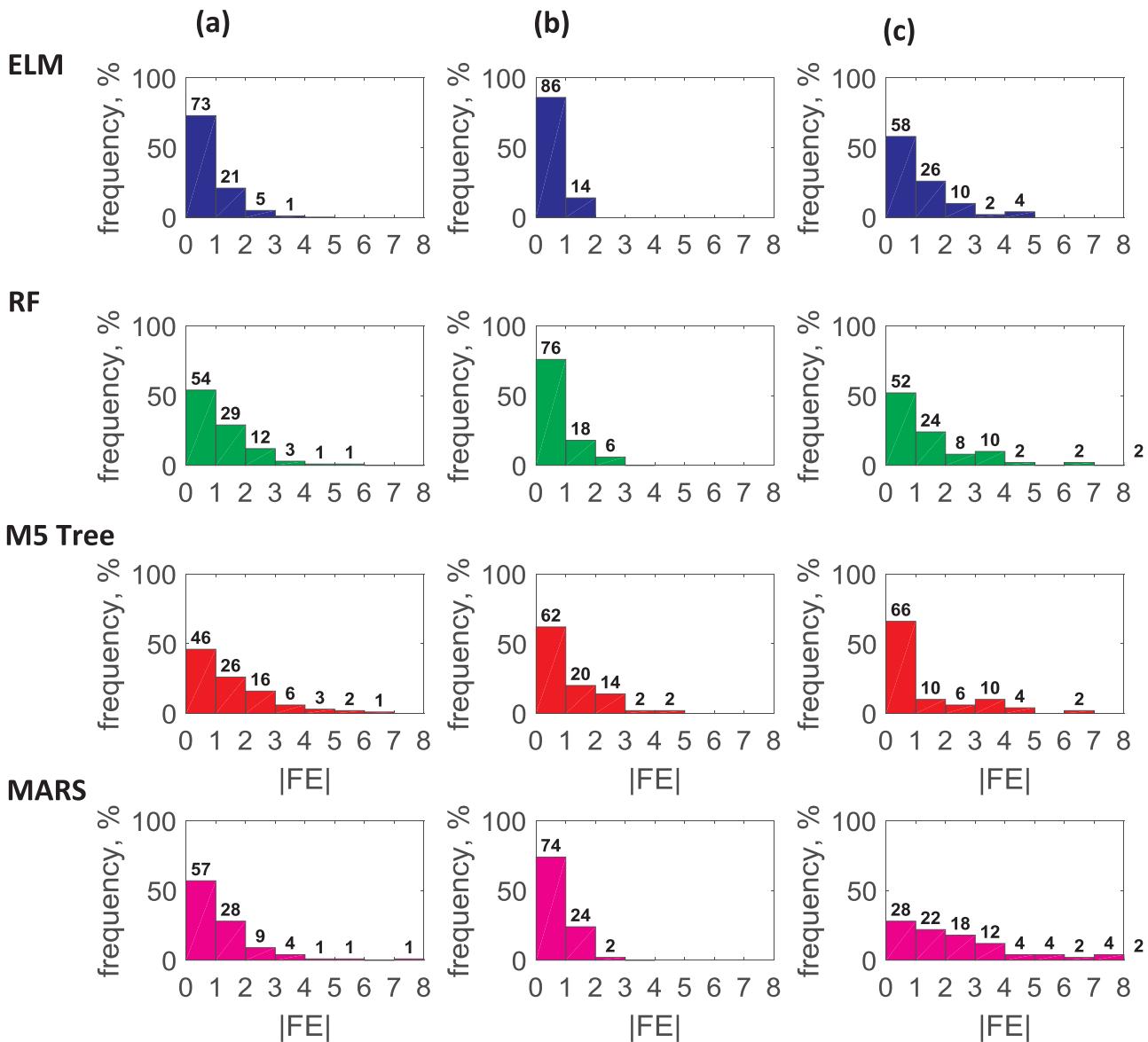


Fig. 12. Frequency plot of absolute forecasted error, $|FE|$ generated by ELM vs. the RF, M5 Model Tree, and MARS models in the test phase. (a) An average of the 15 test sites, (b) most accurately forecasted station (*i.e.*, ST6, White Cliffs), (c) least accurately forecasted station (*i.e.*, ST10, Lake Evella). The numbers denoted on each histogram show the approximate percentage of the forecasted error in the respective error bracket and each horizontal bin shows the cumulative frequency of the forecasted error in $\pm 1 \text{ MJ m}^{-2} \text{ day}^{-1}$ error bracket.

In Fig. 12(a)-(c) a frequency plot of absolute forecasted error $|FE|$ is shown, averaged for all 15 study sites, and for the most and the least accurate study sites. The ELM model registers about 73% of all forecasted errors in the smallest error bracket ($\pm 1 \text{ MJ m}^{-2} \text{ day}^{-1}$) compared to about 54%, 46%, and 57% for the RF, M5 Tree, and MARS models, respectively, for an average of all 15 tested study sites. Similarly, the least accurate station (ST10, Lake Evella) shows approximately 58% (ELM) vs. 52% (RF), 66% (M5 Tree), and 28% (MARS) for the $\pm 1 \text{ MJ m}^{-2} \text{ day}^{-1}$ error bracket. However, for the most accurate study site (ST6, White Cliffs), the performance of ELM and RF models is more distinct with 88% vs. 76%, respectively, of all forecasted errors within the $\pm 1 \text{ MJ m}^{-2} \text{ day}^{-1}$ (Fig. 12b).

To investigate monthly variations in forecasted errors, Fig. 13 displays a polar plot with the error averaged for 15 tested study sites. The spread of these errors across the radial axis is significantly smaller for the ELM compared to the other models, but the largest forecasted error occurs in January for all predictive models. The most accurate performance for the ELM model is recorded between April and October, with

the forecasted errors being lower than $4 \text{ MJ m}^{-2} \text{ day}^{-1}$.

Boxplots displaying the distribution of forecasted error (FE) are shown in Fig. 14. To make a fair judgment only the worst performing study site (ST 6, White Cliffs) (Fig. 14(a)) and best performing study sites (ST 10, Lake Evella) (Fig. 14(b)) are exemplified where the outliers show extreme FE . As with previous results (Figs. 11 and 12; Table 3), the ELM model applied at both study sites registers low forecast errors, generally low upper and low quartile values, and outliers (or extreme errors) located within a smaller range than the RF, M5 Tree, and MARS models. In fact, for the study site where ISR is forecasted most accurately (White Cliffs), the horizontal extent of the box is much narrower and largely centred within the vicinity of $FE = 0$, with outliers extending between $-1.93 \leq FE \leq 1.88 \text{ MJ m}^{-2} \text{ day}^{-1}$. By contrast, the errors for RF extended from $-2.64 \leq FE \leq 2.05 \text{ MJ m}^{-2} \text{ day}^{-1}$, M5 Tree from $-3.32 \leq FE \leq 4.09 \text{ MJ m}^{-2} \text{ day}^{-1}$, and MARS from $-2.31 \leq FE \leq 1.96 \text{ MJ m}^{-2} \text{ day}^{-1}$ (Fig. 14(a)). Interestingly, with forecasted ISR at the least accurate study site (Fig. 14(b)), the performance of both ELM and RF are similar and far exceed that of the M5

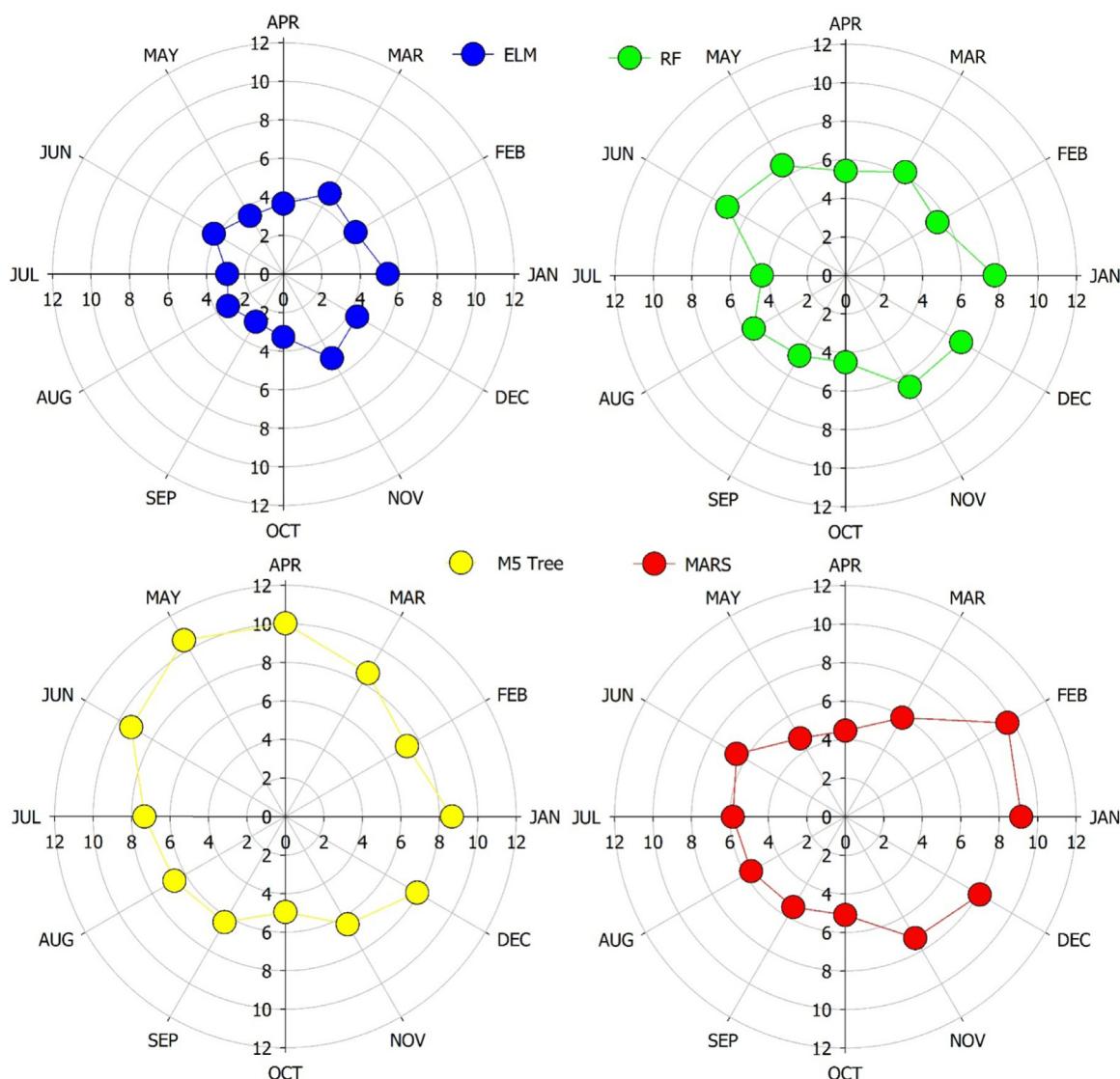


Fig. 13. Relative forecasted error (%) generated by the ELM vs. the RF, M5 Tree, and MARS models over the monthly period in the testing phase for all 15 stations with data averaged over all tested sites considered in Australia. The radial axis outward from the origin denotes the magnitude of the forecasted error for each month.

Tree and MARS model. The MARS model generates an outlier error that is twice that of the ELM and RF models. The high performance of the ELM model is also evident for the all-site average and the individual test sites (not shown here).

5. Concluding remarks

In remote locations such as in the Australian desert environment, big data analytics and emerging computational technologies that use remotely-sensed predictors developed for solar radiation forecasting over long-term periods, are an attractive stratagem for decision-makers to promote solar energy use. These locations do not have environmental monitoring measurement infrastructure and are reliant on predictor data outsourced from satellites or atmospheric reanalysis. Data from MODIS Terra (a multi-national NASA scientific satellite in a sun-synchronous orbit), sweeping a 2,330-km-wide viewing swath and seeing every point on the global scale every 1–2 days in 36 discrete spectral bands, can provide useful modelling information on potential atmospheric variables that may be used to explore future solar radiation on earth. The satellite-based model developed in this study could be a particularly useful decision-support tool for energy utilisation in data sparse regions, and could help facilitate core decisions about future

sustainability of solar energy investments in metro, regional, and remote locations.

In this study, for the first time, a universally-trained extreme learning machine (ELM) model has been designed with eight MODIS satellite data variables, supplemented with geo-topographical site characteristics. The study incorporates a limited temporal dataset spanning 2012–2015 using 26 sites spread evenly across Australia to design the predictive model, and a further 15 sites to evaluate performance. While the MODIS Terra sensor is able to provide a large suite of variables, only the ones most likely to modulate ground-based solar radiation were applied. The ELM model's performance was benchmarked with random forest (RF), M5 Tree, and multivariate adaptive regression splines (MARS) using statistical score metrics and a forecasted data visualization approach. The high performance of ELM in the testing phase was confirmed by relative root mean square errors (less than 6%) obtained for the 15 tested study sites (Fig. 8), which are significantly lower than those obtained by the RF, M5 Tree, and MARS models. Using Legate & McCabe's Index (Fig. 10), the ELM model had an index exceeding 0.50 for 12 out of the 15 study sites, which concurs with the relatively large Willmott's Index and Nash Sutcliffe's coefficient. In terms of a persistence-type model, ELM had a relatively large skill score metric ($RMSE_{SS}$) exceeding all comparison models at all

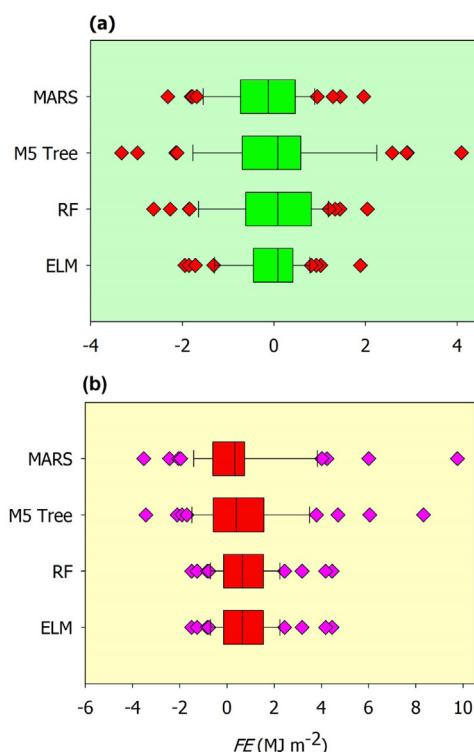


Fig. 14. Boxplots showing the actual error generated between the forecasted and observed ISR ($FE = ISR^{for} - ISR^{obs}$) by the ELM vs. the RF, M5 Tree, and MARS models in the testing phase for: (a) most accurately forecasted station (ST6, White Cliff), (b) least accurately forecasted station (ST10, Lake Evelia).

tested sites, with a site-average value of $RMSE_{SS} = 0.673$ vs. 0.544, 0.408 and 0.298, for the RF, M5 Tree, and MARS models, respectively. Taylor diagrams and error distributions confirmed the superiority of ELM against the three benchmark models explored in this study, demonstrating the utility of ELM for large spatial-scale prediction of global solar radiation in Australia, and perhaps elsewhere, upon further testing of the predictive model.

While the usefulness of ELM was established in this study, and while its application over much of Australia seems to improve the scope of previous studies (e.g., [7,23,27]), there are some limitations that require further consideration. The purpose of this study was to demonstrate the capability of the ELM model for monthly-scale forecasting of global solar radiation, which was largely constrained by the unavailability of daily or shorter time-scale data (given the 1–2 day span of MODIS Terra satellite). However, for real-time predictions and for day-to-day energy sustainability decisions where one may require a predictive model for daily, hourly, and even sub-hourly forecasts, the 6-hourly reanalysis data from the Interim-ERA European Centre for Medium-Range Weather Forecasts (ECMWF) may be used to simulate global solar radiation over large spatial regions. Further enhancement of the ELM model performance could include the utilisation of multi-resolution analysis (i.e., a data pre-processing scheme) to address the inherent non-stationary behaviours in climate-based predictors. Ancillary frequency resolution tools such as the complete ensemble empirical mode decomposition with adaptive noise and the non-decimated discrete wavelet transform [125,126] could be applied. In addition, the application of a feature selection algorithm into the prescribed ELM model where satellite and reanalysis data could be merged to attain a relatively large pool of initial predictor variables (e.g., [27]) could help secure more relevant forecasting variables.

In spite of some limitations, this study provides the foundation for future work, particularly for data-sparse locations where solar energy feasibility studies are required to expand existing solar energy

investments and to support National Electricity Markets. This work is beneficial not only in Australia but also globally, to address climate change issues, devise new energy modelling technologies, and promote sustainable energy resources as per the United Nations Development Program Goal 7. Identification of remotely-located sites through a universally-trained model can support the case for new energy technology financing and bolder energy policy commitments related to solar powered investments, together with promoting the willingness of large nations to embrace technologies on a much wider spatial scale, as promulgated by UNDP initiatives on sustainable energy. Finally, an ELM-based approach utilizing satellite predictor variables, as proposed in this paper, could be further developed as a useful decision-support tool to estimate incident solar radiation (*ISR*) more precisely in locations without proper *ISR* infrastructure. This approach could also aid in solar plant design, the examination of optimal localizations, and could help support new measures for negating electrical power transmission problems through the utilisation of renewable energy.

Acknowledgments

Data were acquired from the Australian Bureau of Meteorology, SILO database, and the MODIS Terra satellite data sensors from the National Aeronautics and Space Administration, all of which are duly acknowledged. The authors are also grateful to the Editor-in-Chief (Dr Aoife M Foley) and both reviewers for their expertise, rigorous reviews, and time spent reviewing the original manuscript.

References

- [1] UNDP. UNDP Strategic Plan, 2018–2021. Special Session (Executive Board of the United Nations Development Programme, the United Nations Population Fund and the United Nations Office for Project Services). New York: United Nations; 2017.
- [2] APVI. IEA Releases the 2016 International Trends in Photovoltaics Report (Australian PV Institute) [cited 2018]; 2016.
- [3] REN. Renewables 2016 Global Status Report: ISBN 978-3-9818107-0-7, Renewable Energy Policy Network for the 21st Century: REN21 Secretariat: Paris; 2016.
- [4] Martin NJ, Rice JL. Developing renewable energy supply in Queensland, Australia: a study of the barriers, targets, policies and actions. Renew Energy 2012;44:119–27.
- [5] REEAR. Renewable Energy (Electricity) Amendment (Exemptions and Other Measures) Regulations, Federal Government: Federal Government, Canberra. Online <<https://www.legislation.gov.au/Details/F2017L01639>>; 2017.
- [6] ARERA. Australian energy resource assessment. Canberra: Commonwealth of Australia (Geoscience Australia) 2014; 2014.
- [7] Deo R, C, Sahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. Renew Sustain Energy Rev 2017;72:828–48.
- [8] Davy RJ, Huang JR, Troccoli A. Improving the accuracy of hourly satellite-derived solar irradiance by combining with dynamically downscaled estimates using generalised additive models. Sol Energy 2016;135:854–63.
- [9] Janjai S, Wattan R, Sripradit A. Modeling the ratio of photosynthetically active radiation to broadband global solar radiation using ground and satellite-based data in the tropics. Adv Space Res 2015;56(11):2356–64.
- [10] Sahin M, et al. Application of extreme learning machine for estimating solar radiation from satellite data. Int J Energy Res 2014;38(2):205–12.
- [11] Linares-Rodriguez A, et al. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 2013;61:636–45.
- [12] Rahimikhooh A, Behbahani S, Banihabib M. Comparative study of statistical and artificial neural network's methodologies for deriving global solar radiation from NOAA satellite images. Int J Climatol 2013;33(2):480–6.
- [13] Blanksby C, Bennett D, Langford S. Improvement to an existing satellite data set in support of an Australia solar atlas. Sol Energy 2013;98:111–24.
- [14] Sharma NP, et al. Satellite estimation and ground measurements of solar UV index and influence of aerosol in ground based UV data in Kathmandu, Nepal. Int J Environ Sci 2013;3(5):1362.
- [15] Senkal O, Kuleli T. Estimation of solar radiation over Turkey using artificial neural network and satellite data. Appl Energy 2009;86(7):1222–8.
- [16] Perez R, Seals R, Zelenka A. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. Sol Energy 1997;60(2):89–96.
- [17] Escobar RA, et al. Estimating the potential for solar energy utilization in Chile by satellite-derived data and ground station measurements. Sol Energy 2015;121:139–51.
- [18] Pfenniger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016;114:1251–65.
- [19] Aguiar LM, et al. Use of satellite data to improve solar radiation forecasting with

- Bayesian Artificial Neural Networks. *Sol Energy* 2015;122:1309–24.
- [20] Boland J, Scott L, Luther M. Modelling the diffuse fraction of global solar radiation on a horizontal surface. *Environmetrics* 2001;12(2):103–16.
- [21] Boland J, Ridley B, Brown B. Models of diffuse solar radiation. *Renew Energy* 2008;33(4):575–84.
- [22] Huang J, et al. Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model. *Sol Energy* 2013;87:136–49.
- [23] Deo RC, Wen X, Feng Q. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 2016;168:568–93.
- [24] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. *Neurocomputing* 2006;70(1):489–501.
- [25] Deo RC, Şahin M. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environ Monit Assess* 2016;188(2):90.
- [26] Deo R, C, Şahin M. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmos Res* 2015;153:512–25.
- [27] Salcedo-Sanz S, et al. An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Appl Energy* 2018;209:79–94.
- [28] Currie JJ, Goulet PJ, Ratsimandresy AW. Wind conditions in a fjordlike bay and predictions of wind speed using neighboring stations employing neural network models. *J Appl Meteorol Climatol* 2014;53(6):1525–37.
- [29] Bilgili M, Sahin B, Yasar A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renew Energy* 2007;32(14):2350–60.
- [30] Damousis IG, et al. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Trans Energy Convers* 2004;19(2):352–61.
- [31] Deo RC, et al. Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data. *Renew Rev Energy* 2018;116(A):309–23.
- [32] Velázquez S, Carta JA, Matías J. Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: a case study in the Canary Islands. *Appl Energy* 2011;88(11):3869–81.
- [33] Casella L. Improving long-term wind speed assessment using joint probability functions applied to three wind data sets. *Wind Eng* 2012;36(4):473–83.
- [34] Şahin M, Kaya Y, Uyar M. Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data. *Adv Space Res* 2013;51(5):891–904.
- [35] Şahin M. Modelling of air temperature using remote sensing and artificial neural network in Turkey. *Adv Space Res* 2012;50(7):973–85.
- [36] Saadat H, et al. Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS J Photogramm Remote Sens* 2011;66(5):608–19.
- [37] Marj AF, Meijerink AM. Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *Int J Remote Sens* 2011;32(24):9707–19.
- [38] Şenkal O. Modeling of solar radiation using remote sensing and artificial neural network in Turkey. *Energy* 2010;35(12):4795–801.
- [39] Carter D, Curran PJ. Principles of Remote Sensing, JSTOR; 1985.
- [40] Pelton JN, Madry S, Camacho-Lara S. Handbook of satellite applications. Incorporated: Springer Publishing Company; 2012.
- [41] Price JC. Land surface temperature measurements from the split window channels of the NOAA 7 Advanced Very High Resolution Radiometer. *J Geophys Res: Atmos* 1984;89(D5):7231–7.
- [42] Becker F, Li Z-L. Towards a local split window method over land surfaces. *Remote Sens* 1990;11(3):369–93.
- [43] Olivieri C, et al. A split window algorithm for estimating land surface temperature from satellites. *Adv Space Res* 1994;14(3):59–65.
- [44] Kerenyi J, Putsay M. Investigation of land surface temperature algorithms using NOAA AVHRR images. *Adv Space Res* 2000;26(7):1077–80.
- [45] MODIS. MODIS (Moderate-Resolution Imaging Spectroradiometer, NASA, Editor, NASA: In the Internet; 2016. <http://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf>.
- [46] Benali A, et al. Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sens Environ* 2012;124:108–21.
- [47] Qin J, et al. Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. *Appl Energy* 2011;88(7):2480–9.
- [48] Huete A, Justice C, Leeuwen Wv. Modis vegetation index (MOD 13) algorithm theoretical basis document. USA: University of Arizona, University of Virginia and NASA; 1999.
- [49] Wan Z, Dozier J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans Geosci Remote Sens* 1996;34(4):892–905.
- [50] Wan Z, et al. Quality assessment and validation of the MODIS global land surface temperature. *Int J Remote Sens* 2004;25(1):261–74.
- [51] Coll C, et al. Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data. *Remote Sens Environ* 2005;97(3):288–300.
- [52] Wan Z. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD). *Inst Comput Earth Syst Sci St Barbar* 1999:75.
- [53] Menzel P, Strabala K. Cloud top properties and cloud phase algorithm theoretical basis document. University of Wisconsin–Madison; 1997.
- [54] Platnick S, et al. The MODIS cloud optical and microphysical products: collection 6 updates and examples from Terra and Aqua. *IEEE Trans Geosci Remote Sens* 2017;55(1):502–25.
- [55] Ahmad SP, et al. MODIS cloud, aerosol, and water vapor products for climate and global change studies. In: Proceedings of the 13th symposium on global change and climate variations; 2002.
- [56] Weisz E, et al. Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals. *Geophys Res Lett* 2007;34:17.
- [57] Menzel WP, et al. MODIS global cloud-top pressure and amount estimation: algorithm description and results. *J Appl Meteorol Climatol* 2008;47(4):1175–98.
- [58] Chahine MT. Remote sounding of cloudy atmospheres. I. The single cloud layer. *J Atmos Sci* 1974;31(1):233–43.
- [59] Smith W, Platt C. Comparison of satellite-deduced cloud heights with indications from radiosonde and ground-based laser measurements. *J Appl Meteorol* 1978;17(12):1796–802.
- [60] Wylie DP, Menzel W. Two years of cloud cover statistics using VAS. *J Clim* 1989;2(4):380–92.
- [61] Naud C, Müller JP, Clothiaux EE. Comparison of cloud top heights derived from MISR stereo and MODIS CO₂-slicing. *Geophys Res Lett* 2002;29(16).
- [62] Baum BA, et al. MODIS cloud-top property refinements for Collection 6. *J Appl Meteorol Climatol* 2012;51(6):1145–63.
- [63] Iwabuchi H, et al. Retrieval of radiative and microphysical properties of clouds from multispectral infrared measurements. *Prog Earth Planet Sci* 2016;3(1):32.
- [64] Vijayakumar K, Devara P. Study of aerosol optical depth, ozone, and precipitable water vapour content over Sinhagad, a high-altitude station in the Western Ghats. *Int J Remote Sens* 2013;34(2):613–30.
- [65] Seemann SW, et al. Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances. *J Appl Meteorol* 2003;42(8):1072–91.
- [66] Borbas E, et al. MODIS atmospheric profile retrieval algorithm theoretical basis document. Electronic resource; 2011. <http://modis-atmos.gsfc.nasa.gov/MOD07_L2/atbd.html>, (10 August 2015).
- [67] Huang GB, Chen L, Siew CK. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Netw* 2006;17(4):879–92.
- [68] Huang G, et al. Trends in extreme learning machines: a review. *Neural Netw* 2015;61:32–48.
- [69] Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- [70] Cutler DR, et al. Random forests for classification in ecology. *Ecology* 2007;88(11):2783–92.
- [71] Prasad AM, Iverson LR, Liaw A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 2006;9(2):181–99.
- [72] Bylander T. Estimating generalization error on two-class datasets using out-of-bag estimates. *Mach Learn* 2002;48(1–3):287–97.
- [73] Quinlan JR. Learning with continuous classes. In: Proceedings of the 5th Australian joint conference on artificial intelligence. Singapore; 1992.
- [74] Mitchell TM. Machine learning. Computer Science Series. Burr Ridge: McGraw-Hill; 1997. [MATH, 1997].
- [75] Rahimikhoob A, Asadi M, Mashal M. A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region. *Water Resour Manag* 2013;27(14):4815–26.
- [76] Bhattacharya B, Solomatine DP. Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* 2005;63:381–96.
- [77] Kisi O. Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J Hydrol* 2015;528:312–20.
- [78] Witten IH, Frank E. Data mining: practical machine learning tools and techniques. Morgan Kaufmann; 2005.
- [79] Friedman JH. Multivariate adaptive regression splines. *Ann Stat* 1991;1:67.
- [80] Cheng M-Y, Cao M-T. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Appl Soft Comput* 2014;22:178–88.
- [81] Deo RC, Samui P, Kim D. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stoch Environ Res Risk Assess* 2016;30(6):1769–84.
- [82] Krzyścien J. Nonlinear (MARS) modeling of long-term variations of surface UV-B radiation as revealed from the analysis of Belsk, Poland data for the period 1976–2000. *Annal Geophys* 2003. (Copernicus GmbH).
- [83] Butte NF, et al. Validation of cross-sectional time series and multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents using doubly labeled water. *J Nutr* 2010;140(8):1516–23.
- [84] Sephton P. Forecasting recessions: can we do better on mars. *Federal Reserve Bank of St. Louis Review*. 83(March/April 2001); 2001.
- [85] Zhang W, Goh A. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Comput Geotech* 2013;48:82–95.
- [86] Sharda V, et al. Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptives (MARS) pour la prévision d'écoulement au sein de micro-bassins versants Himalayens d'altitudes intermédiaires avec peu de données. *Hydrol Sci J* 2008;53(6):1165–75.
- [87] Craven P, Wahba G. Smoothing noisy data with spline functions. *Numer Math* 1978;31(4):377–403.
- [88] Milborrow S. Multivariate Adaptive Regression Splines. Package ‘earth’: Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller’s Fortran utilities with Thomas Lumley’s leaps wrapper. UR; 2016. <<http://www.milbo.users.sonic.net/earth>>.
- [89] Hastie T, Tibshirani R, Friedman J. Unsupervised learning. The elements of

- statistical learning. Springer; 2009. p. 485–585.
- [90] Jekabsons G. Adaptive regression splines toolbox for Matlab/Octave. Version 2013;1:72.
- [91] ABARE. Australian energy projections to 2029–2030. Australian Bureau of Agricultural and Resource Economics; 2010.
- [92] AEU, Australian Energy Update. Department of industry and science (2015). Canberra, August: Australian Energy Update; 2015.
- [93] Raupach M, et al. Australian water availability project (AWAP): CSIRO marine and atmospheric research component: final report for phase 3. Melbourne, Australia: Centre for Australian weather and climate research (bureau of meteorology and CSIRO); 2009.
- [94] AWAP. Readme File: Australian Water Availability Project (AWAP), CSIRO: CSIRO; 2016.
- [95] WG and MMA. High temperature solar thermal technology roadmap in Wyld Group and MMA (McLennan Magasanik Associates). Australia: Prepared for: New South Wales and Victorian Governments Victoria; 2008.
- [96] Jeffrey SJ, et al. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ Model Softw 2001;16(4):309–30.
- [97] BOM. Australian Data Archive for Meteorology for Meteorology; 2015.
- [98] Beesley C, Frost A, Zajaczkowski J. A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets. Australia: 18th World IMACS/MODSIM Congress, Cairns; 2009.
- [99] Tozer C, Kiem A, Verdon-Kidd D. On the uncertainties associated with using gridded rainfall data as a proxy for observed. Hydrol Earth Syst Sci 2012;16(5):1481–99.
- [100] Zajaczkowski J, Wong K, Carter J. Improved historical solar radiation gridded data for Australia. Environ Model Softw 2013;49:64–77.
- [101] Deo R, C, Kisi O, Singh VP. Drought forecasting in eastern Australia using multi-variate adaptive regression spline, least square support vector machine and M5Tree model. Atmos Res 2017;184:149–75.
- [102] Barzegar R, et al. Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. Sci Total Environ 2018;621:697–712.
- [103] Deo R, C, et al. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. Environ Res 2017;155:141–66.
- [104] Jekabsons G. M5PrimeLab M5': regression tree and model tree toolbox for Matlab/Octave. Reference manual for M5Prime. ver. 1.0.1; 2010.
- [105] Wang YW. IH: Inducing Model Trees for Predicting Continuous Classes. In: Proceedings of European Conference on Machine Learning. University of Economics, Prague; 1997.
- [106] Kooperberg C, Clarkson DB. Hazard regression with interval-censored data. Biometrics 1997;1485–94.
- [107] Zareipour H, Bhattacharya K, Canizares C. Forecasting the hourly Ontario energy price by multivariate adaptive regression splines. In: Proceedings of 2006 IEEE Power Engineering Society General Meeting. IEEE; 2006.
- [108] ASCE. Criteria for evaluation of watershed models. J Irrig Drain Eng 1993;119(3):429–42.
- [109] Nash J, Sutcliffe J. River flow forecasting through conceptual models part I—A discussion of principles. J Hydrol 1970;10(3):282–90.
- [110] Willmott CJ. Some comments on the evaluation of model performance. Bull Am Meteorol Soc 1982;63(11):1309–13.
- [111] Dawson CW, Abrahart RJ, See LM. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environ Model Softw 2007;22(7):1034–52.
- [112] Legates DR, McCabe GJ. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. Water Resour Res 1999;35(1):233–41.
- [113] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. Geosci Model Dev 2014;7(3):1247–50.
- [114] Legates DR, Davis RE. The continuing search for an anthropogenic climate change signal: limitations of correlation-based approaches. Geophys Res Lett 1997;24(18):2319–22.
- [115] Willmott CJ. On the validation of models. Phys Geogr 1981;2(2):184–94.
- [116] Willmott CJ, Matsura K, Robeson SM. Ambiguities inherent in sums-of-squares-based error statistics. Atmos Environ 2009;43(3):749–52.
- [117] Willmott CJ. On the evaluation of model performance in physical geography. Spatial statistics and models. Springer; 1984. p. 443–60.
- [118] Wilcox BP, et al. Predicting runoff from rangeland catchments: a comparison of two models. Water Resour Res 1990;26(10):2401–10.
- [119] Garrick M, Cunnane C, Nash J. A criterion of efficiency for rainfall-runoff models. J Hydrol 1978;36(3–4):375–81.
- [120] Mohammadi K, et al. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manag 2015;92:162–71.
- [121] Ertekin C, Yıldız O. Comparison of some existing models for estimating global solar radiation for Antalya (Turkey). Energy Convers Manag 2000;41(4):311–30.
- [122] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmos 2001;106(D7):7183–92.
- [123] Mohammadi K, et al. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Convers Manag 2015;91:433–41.
- [124] Iverson L, Prasad A, Liaw A. New machine learning tools for predictive vegetation mapping after climate change: bagging and random forest perform better than regression tree analysis. In: Proceedings of UK-International Association for Landscape Ecology. Cirencester, UK; 2004.
- [125] Al-Musaylh MS, et al. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. Appl Energy 2018;17:422–39.
- [126] Prasad R, et al. Input selection and performance optimization of ANN-based streamflow forecasts in a drought-prone Murray Darling Basin using IIS and MODWT algorithm. Atmos Res 2017;197:42–63.